MOLECULAR SIZE-BASED MODEL TO ESTIMATE PHARMACEUTICALLY RELEVANT PHYSICOCHEMICAL PROPERTIES

By

PETER BUCHWALD

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1997

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. N. Bodor for the opportunity to work in his center on such a uniquely broad spectrum of diverse areas and for his excellent guidance, inspiring input, and constant encouragement.

I gratefully appreciate the help received from Dr. M. D. C. Barros, Dr. L. Buris, Dr. H. Farag, Dr. A. Harget, Dr. M.-J. Huang, Dr. A. Juhasz, Dr. K. Prokai-Tatrai, Dr. L. Prokai, Dr. G. Somogyi, Dr. J.-X. Wu, and Dr. W.-M. Wu. I would also like to extend my thanks to members of the Center, J. Berger, K. Eberst, E. Guy, L. Johnston, J. Martignano, and R. Wong, for their helpful assistance, as well as to all the friends and colleagues who made this tenure pleasant.

Finally, I would like to thank the members of my supervisory committee, Drs. H. Derendorf, W. J. Millard, and D. A. Micha, for their interest in the project.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS ii

KEY TO SYMBOLS iii

ABSTRACT vii

CHAPTERS

1 INTRODUCTION 1

2 MOLECULAR SIZE 3

Background 3
Algorithm 8
Results and Discussion 11

3 UNIFIED MODEL FOR LIQUID PROPERTIES 15

Background—Statistical Mechanics 16
Experimental Data 24
Enthalpy of Vaporization 25
Boiling Point 27
Gas-Liquid Solubility and Vapor-Liquid Equilibrium 30
Partition Between Two Solvents 35
Solubility (Liquid-Liquid) 37

4 OCTANOL-WATER PARTITION 40

Background 40
Experimental Data 50
Model—Results and Discussion 51
Model for Log $P_{o/w}$ 51
Hydrogen Bonding? 64
Comparison with Other Methods 67

5 NONZWITTERIONIC PEPTIDES: OCTANOL-WATER PARTITION 75

Background 75
Experimental Data 78
Model 80
Partition Coefficients 82

Apparent Partition Coefficients (Monoionic Species) 91
Cyclic Peptides 98
Cyclosporin A 99
Antitumor Hexapeptides RA 100
Gramicidin S Analogs 100
Desmopressin Derivatives 101
Melanotan-II 102
Sandostatin (Octreotide) 102

6 RETROMETABOLIC DRUG DESIGN: PHYSICOCHEMICAL CONSIDERATIONS 104

Background 104
Experimental Determinations 106
Principles of Retrometabolic Drug Design 109
Chemical Delivery Systems 111
Site-Specific Enzyme-Activated Chemical Delivery System 113
Design Principles 113
Physicochemical Data—Calculations and Measurements 116
Enzymatic Physical-Chemical–Based Chemical Delivery Systems 123
Design Principles 123
Physicochemical Data—Calculations and Measurements 129

7 CONCLUSIONS 135

REFERENCES 138

BIOGRAPHICAL SKETCH 158

KEY TO SYMBOLS

α - angle between atomic centers

β - hydrogen bond acceptor basicity (solvatochromic model)

φ - half-angle under which an intersection circle is seen from the center of a sphere

μ - chemical potential

ω, ω_0, ω_{ij} - constants characterizing the W interaction potential ($\omega = w/kT_0$)

ρ - number density (N/V)

γ - Ostwald absorption coefficient

σ - standard error

Λ - thermal de Broglie wavelength

κ_g - geodesic curvature

A - Helmholtz free energy

$\mathcal{A}$ - surface area

a, b - van der Waals constants for gases

c - molar concentration (n_i/V)

D - distribution (apparent partition) coefficient

d_{ij} - interatomic distance

E - energy

E_c - extraction coefficient

G - Gibbs free energy

H - enthalpy of vaporization

h - Planck constant (6.6262×10^{-34} J s)

K_a - ionic equilibrium constant

K_e - extraction constant

K_p - ion pair association constant

k - Boltzmann constant (1.3807×10^{-23} J K^{-1})

k' - chromatographic capacity factor

m - mass

N - number of molecules

N_0 - Avogadro's number (6.0220×10^{23} mol^{-1})

$\mathcal{N}$ - new parameter introduced in log $P_{o/w}$ model

n - number of data points

n_i - number of moles for species i

P - partition coefficient

P_p - (true) partition coefficient of ion pairs

p - pressure

Q - canonical ensemble partition function

q - molecular partition function

r_i - radius of sphere i

r - correlation coefficient

R - universal gas constant (8.3144 J K^{-1} mol^{-1})

S - entropy

T - absolute temperature

T_o - room temperature (298.15 K)

T_b - boiling point

v - molecular volume

$\mathcal{V}$, V - volume

V^0 - molar volume

W - interaction potential

w, w_0, w_i - constants characterizing the W interaction potential

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MOLECULAR SIZE-BASED MODEL TO ESTIMATE PHARMACEUTICALLY RELEVANT PHYSICOCHEMICAL PROPERTIES

By

Peter Buchwald

December, 1997

Chairman: Nicholas S. Bodor
Major Department: Pharmaceutics

The present work explores the possibilities of a unified, molecular size-based model to predict pharmaceutically relevant physicochemical properties from molecular structure. The main focus is on octanol-water partition with the following linked, but separate subjects covered: (1) A new algorithm to estimate van der Waals molecular volume and surface area combines exact analytical methods with fast numerical techniques to obtain sufficiently fast and reliable intrinsic size estimates that can be useful in QSAR and other studies. (2) A statistical mechanics based model allows unified description of enthalpies of vaporization, boiling points, partition coefficients, and solubilities for organic liquids having no strongly polar or hydrogen bonding substituents. Molecules in a liquid are considered as moving freely in a fraction of the total volume within

an average attractive potential described by molecular volume through a linear relationship. The obtained equations, relying on computed molecular volume alone, account for 90-95% of the variance of these properties. (3) A remarkably simple model predicts log octanol-water partition coefficients (log*P*). It is obtained from the previous model by introducing a new, quantified parameter determined by the functional groups of the molecule. This parameter correlates well with hydrogen bond acceptor basicity, and the corresponding free energy change agrees well with that accepted for hydrogen bonding in water. Tested on over 700 organic compounds, this distinctively simple and fully computerized 3D approach compares favorably with empirical, complicated, "two-dimensional" fragment methods. (4) The above log*P* model is applied for nonzwitterionic peptides. Promising correlations are obtained, and predictions are realistic even for larger peptides where common, overparametrized fragment methods completely fail. Ion-pair partitioning is reviewed, and estimation of distribution coefficients measured around physiological conditions for monoionic peptides is attempted. (5) Physicochemical considerations related to retrometabolic drug design approaches are discussed. Describing the partition of quaternary pyridinium-type compounds, the present model suggests a change of about five orders of magnitude compared with the corresponding neutral forms. Such significant change explains the "lock-in" mechanism of brain-targeting chemical delivery systems using a targetor moiety of this type. Experimental considerations and some newly determined values are also included.

CHAPTER 1
INTRODUCTION

Reliable prediction of physicochemical or biological properties from chemical structure alone is probably the ultimate goal in computer-aided drug design. The hardware and software development of the last years made possible significant advances, but reliable and general models are still far away. Most applications, especially those in pharmaceutical fields, still rely on less accurate, one- or two-dimensional size estimates, such as number of carbon atoms, molecular weight, number of different fragments, and so on. The present work explores the possibilities of a unified, statistical mechanics-based model that uses computed 3D molecular size to predict properties for liquid organic compounds. This approach allows a unified treatment with only a limited number of parameters.

The main subject is the octanol-water partition coefficient, a successful parameter widely used to characterize molecular lipophilicity, but the included chapters cover a full circle from the computer algorithm implemented for size computation through the developed predictive model to concrete applications in the field of retrometabolic drug design. They present, therefore, the following five linked, but more or less separate subjects:

• A novel, sufficiently fast and accurate computational algorithm combines analytical and numerical techniques to obtain van der Waals molecular volume and surface area for molecules with given 3D structure.

• A possible model that relies on molecular volume to account for non-specific interactions in organic liquids gives a unified, reasonably good description of boiling points, enthalpies of vaporization, and partition properties for molecules having no strongly polar or hydrogen-bonding substituents.

• A fully computerized model, derived from the previous one, predicts octanol-water partition coefficients with a remarkable reduction in the number of parameters used. It was tested on over 700 organic compounds comprising various pharmaceutically useful molecules, including even nucleosides and peptides.

• The above model is applied for a large database of nonzwitterionic peptides that includes some large compounds, such as cyclosporin A, octreotide, or gramicidin. Problems related to the partition of ionic species are also discussed here.

• Finally, physicochemical considerations that exploit novel possibilities offered by the model and are related to retrometabolic drug design approaches are discussed. Experimental considerations, together with a few newly determined values, are also included within this last chapter.

CHAPTER 2
MOLECULAR SIZE

Background

Obtaining the accurate molecular volume or surface area of a molecule with a given spatial structure is an important objective as most physicochemical properties (solubility, partition coefficient, boiling point, etc.) and many biological properties are strongly size related. Most applications, especially those in pharmaceutical fields, rely on less accurate, one- or two-dimensional size estimates like number of carbon atoms, molecular weight, number of bonds, etc. A novel algorithm to compute van der Waals molecular volume and surface area is presented here. This algorithm is fast and accurate enough even for large, biologically important molecules and should be useful in different 3D based models. One such model that achieves a remarkable reduction in the number of parameters used to predict partition properties is presented later in chapter 4.

The empirical success of the van der Waals radius concept gives a good starting point for a computational approach, even if in an exact quantum chemical description the electron cloud has no well-defined boundary surface. Each atom of the molecule is represented as a sphere centered at the equilibrium position of the atomic nucleus and having a radius equal to the van der Waals radius of the corresponding atom, the radius being defined in terms of

the distance of closest approach of atoms or molecules that are not bonded (Laidler and Meiser, 1995). The exterior surface of all these spheres defines the van der Waals surface, which delimits the so-called van der Waals volume of the molecule. The solvent accessible surface is defined as the surface traced by the center of the spherical test probe (modeling the solvent molecule) as it is rolled over the van der Waals surface (Lee and Richards, 1971). Sometimes, to avoid the strong solvent dependency of this surface, it is replaced by the contact surface: the smooth network of convex and reentrant surface traced by the inward-facing part of the probe sphere as it rolls over the molecule (Richards, 1977). Bader et al. (1987) have shown that contours of constant electronic density (0.001-0.002 electrons/bohr^3) that contain over 96% of the total electronic charge gave good representations of molecular van der Waals surfaces, and they gave probably even better representations of the smoother contact surfaces.

There were many attempts to reliably estimate these properties starting with group contribution methods (Bondi, 1964) or even some empirical approaches using space-filling models of the solutes. Later, they were replaced by computational models. Several volume and/or surface area computing algorithms were developed using different numerical techniques to divide each sphere into smaller subunits. Many of these algorithms rely on sampling points either using a regular grid [within a rectangular box enclosing the molecule (Gavezzotti, 1983; Leahy, 1986), within separate boxes built around each atom

(Bodor et al., 1989), placed on the spherical surfaces (Shrake and Rupley, 1973; Le Grand and Mertz, 1993), defined by spherical triangles (Silla et al., 1991)] or just randomly placing them with Monte Carlo techniques (Alejandre et al., 1988; Higo and Go, 1989). Other algorithms use slicing planes placed at defined intervals through the whole molecule (Lee and Richards, 1971), or incrementally rotated around an axis for each atom (Pearlman, 1986; Hermann, 1972).

Starting from 1983, analytical methods were developed to obtain the surface area and sometimes the volume for a set of interpenetrating spheres. The main approaches are those developed by Connolly (1983, 1985), Richmond (1984), Gibson and Scheraga (1987,1988), Dodd and Theodorou (1991), but different variations and improvements were also published (Perrot et al., 1992; Eisenhaber and Argos, 1993; Petitjean, 1994; Liotard et al., 1995). Analytical methods require the determination and characterization of all the vertices (points of triple intersection) and arcs (segments of the circle formed at the intersection of two spheres) that are on the accessible (exterior) surface. Most numerical techniques are easily implemented, but at higher sampling densities they become inefficient. Analytical methods are more complex and more accurate, but they can be very sensitive to roundoff errors, and they can become impractical with an increase in the number of sphere-sphere intersections (Bliznyuk and Gready, 1996).

These calculated geometric properties are related to direct experimental measures of molecular size such as molar volume (molecular weight divided by

liquid density) (Leahy, 1986) or co-volume (the *b* parameter in the van der Waals equation of state for a gas) (Gavezzotti, 1983), and they are already successfully used as size estimates in many different models (Gavezzotti, 1983; Leahy, 1986; Bodor et al., 1989; Bodor and Huang, 1992b, 1992c; Perrot et al. 1992; Kamlet et al., 1987). The values obtained depend on the set of van der Waals radii used. Atomic radii typically used in chemical calculations are close to those given originally by Bondi (1964); for example, that used by Bodor et al. (1989) is: H 1.20, C 1.70, N 1.57, O 1.51, F 1.43, Cl 1.75, Br 1.95, I 2.10, S, P 1.80 (all values in Å). Slightly smaller or bigger values (e.g., for carbon between 1.5 and 2.0) are also used (Silla et al., 1991; Richards, 1977). In the following we will use this set and for comparison one obtained with a 10% reduction for most atoms (H 1.08, C 1.53, N 1.40, O 1.36, F 1.29, Cl 1.60, Br 1.83, I 2.05, S 1.70, P 1.75, Si 2.10). Such a reduction is not uncommon in hard-sphere potential models (Iijima et al., 1987; Creamer and Rose, 1994; Wimley et al., 1996) attempting to reproduce sterically allowed protein conformations. The values obtained and used here are practically identical with the effective van der Waals radii that represent the distance of closest approach and describe the atoms' own volume into which other atoms cannot intrude (Motoc and Marshall, 1985). With the second set the relatively small decrease leads to significantly smaller overlaps between adjacent atoms. Even if the values obtained with the two different sets of radii are numerically different, they are so strongly

correlated that any relation between obtained molecular size and other properties will remain practically unaltered.

The reason behind this correlation is probably the linear relationship between molecular volume and surface area noticed by Pearlman (1986) or Meyer (1986) and confirmed here on a large dataset including more than 400 different organic molecules. This is somehow surprising as for true three-dimensional objects the surface area should be proportional with $V^{2/3}$ and not V itself. This linear relationship might be a factor behind the unexpectedly good correlations between differently defined surfaces (Shrake and Rupley, 1973), as described, for example, between solvent accessible surfaces computed with different solvent radii (Hermann, 1972) or between solvent accessible surface and van der Waals surface (Amidon et al., 1975; Valvani et al., 1976). A study by Connolly (1985) also indicated that within certain limits the change in molecular volume is linearly related to the change in radii. Hermann (1997) demonstrated that molecular and accessible surfaces lead to exactly the same results when calculating solvation free energies. Most likely, due to these linear relationships, so many different radii-sets are still in use since the obtained size estimates correlate well enough to not allow a clear distinction (Livingstone et al., 1991).

Because our main interests are in partition properties and because the solvent molecules involved (e.g., water and *n*-octanol) have frequently very different sizes, we will use solvent independent van der Waals volumes and

surface areas. The values used here were computed with a novel algorithm developed by combining both analytical and numerical techniques to obtain a method that when computing van der Waals molecular size, is sufficiently accurate and is significantly faster than an entirely numerical one. Exact, analytical formulas account for the overlaps between atoms bonded to a common neighbor and for the multiple overlaps in the interior of smaller rings. Whenever other, usually small, overlaps are detected, a local numerical technique is used to evaluate their effect on surface and volume.

Algorithm

In the first step, for each atomic sphere the volume (surface) that is not overlapping with a bonded neighbor is summed up. For two such spheres (with radii r_i and r_j), the total volume (surface) is the sum of the two sphere's volume (surface) excluding the overlapping spherical segments with heights h_i and h_j, respectively:

$$\mathcal{V} = [\frac{4}{3}\pi r_i^3 - \frac{\pi}{3}(3r_i - h_i)h_i^2] + [\frac{4}{3}\pi r_j^3 - \frac{\pi}{3}(3r_j - h_j)h_j^2]$$
$$\mathcal{A} = [4\pi r_i^2 - 2\pi r_i h_i] + [4\pi r_j^2 - 2\pi r_j h_j] \qquad (2.1)$$

These formulas can be extended for more than two spheres introducing the necessary analog terms. This gives the correct total result if there are no overlaps between nonbonded atoms or if these overlaps are completely included inside an atomic sphere. In most chemical compounds the equilibrium distances are such that when dealing with intrinsic van der Waals properties, there are some "external" overlaps of nonbonded atoms. However, they are relatively

small and may occur between larger atoms bonded to a common one (see the shaded area in figure 2.1), in the inside region of smaller rings, and sometimes between structurally distant atoms (crowded or strained structures, intramolecular hydrogen bonds, etc.).

For larger atoms bonded to a common one, the area of a convex face delimited by different intersection arcs is obtained from differential geometry's global Gauss-Bonnet formula (do Carmo, 1976) using the facts that the Gaussian curvature of a sphere and the geodesic curvature κ_g of a circle on the surface are constants:

$$\mathcal{A} = r_k^2 \left(2\pi - \sum_e \kappa_g^e l_e - \sum_v \beta_v\right) \qquad (2.2)$$

Here r_k is the radius of the sphere on which the face lies; the first sum is over all edges that make up its boundary, and the second sum is over the vertices at which these edges intersect; l_e is the arc length of edge e; and β_v is the external angle at vertex v (figure 2.1). The geodesic curvatures involved are constants since they refer to portions of the intersection circles; as an example the one formed between spheres k and i has:

$$\kappa_g^{ki} = \frac{1}{r_k \, tg\, \varphi_{ki}} \qquad (2.3)$$

All angles and arc lengths can be obtained from the coordinates and radii of the three spheres involved using basic trigonometric relations and some vector algebra. The corresponding volume is also computed analytically using,

after generalization for unequal radii, a method originally described by Rowlinson (Rowlinson, 1963) to obtain the volume of a triple overlap.

For the inside region of smaller rings, if the molecular structure contains less than seven-membered rings, additional correction is required to account for multiple overlaps in the inside region of the ring. Therefore, the program includes an algorithm that detects all small rings present in the molecular structure. For each atom in one of these rings the area sliced off from its surface by the other members of the same ring is then computed using essentially the formulae presented before. Here, due to the possibility of multiple overlaps, all the external vertices exposed on one or the other side of the ring have to be detected first, and then the necessary arc lengths are computed using the coordinates of the four centers involved. Unfortunately, in this approach the corresponding exact volume cannot be easily obtained, but for slightly touching spheres, as is the case here, the affected volume is much smaller than the affected surface area. Hence, a ring-size dependent correction that is adjusted as a function of the computed surface is used, a method that proved to be sufficient for practically all reasonable molecular structures.

Finally, for structurally distant atoms, if an overlap is detected and none of the above holds, the necessary surface correction is estimated with a numerical algorithm. Points are placed on the surface of the segment of sphere *i* contained within sphere *j* using a system of spherical coordinates around the axis that connects the two centers, and it is checked whether these grid points are

contained inside other spheres or not. The use of spherical coordinates allows placing the test points only within the involved segment, usually a small fraction of the total sphere; thus high numbers of test points are not required, and these routines are sufficiently fast.

The area of the portion of *i*'s surface that is inside *j*, but is not inside other spheres previously considered relative to *i*, has to be subtracted. Its value is obtained in the usual way, multiplying the area of the total segment of height h_i by the fraction of points that lie within *j* only (N_{oij}/N).

$$A_{cij} = 2\pi r_i h_i \frac{N_{oij}}{N} \qquad (2.4)$$

The corresponding volume correction is again generally quite small; it is simply estimated by multiplying the total volume of the segment with the fraction obtained for the surface.

Results and Discussion

All subroutines were separately tested. The geometric accuracy of the final version was verified by comparing its output with that of an entirely numerical procedure (Bodor et al., 1989) using a set of 100 molecules ranging from methane to steroids and tripeptides. Half the molecules had a simple molecular mechanics optimized structure (Minimize procedure from Alchemy II program; Tripos Assoc., St. Louis, MO), half had a structure obtained with full geometry optimization using the semiempirical AM1 method (Dewar, AMPAC 1.0).

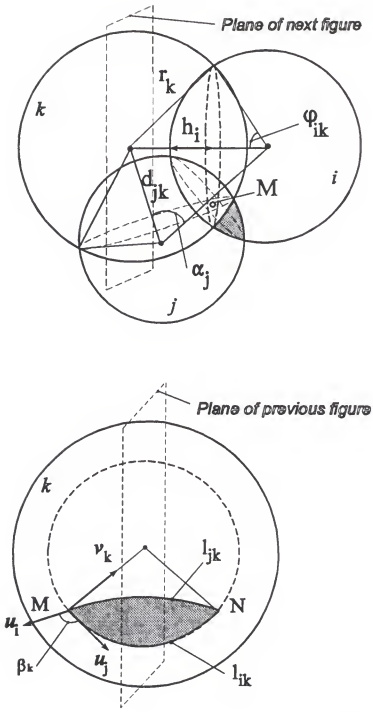


Figure 2.1 A possible overlap of three spheres shown from two different angles. Vertices M and N are points of triple overlap; the dotted ellipses in the first figure represent intersection circles; the shaded area in the second figure is on the surface of sphere *k*. See text for additional details.

The higher the grid density of the numerical program the smaller the average difference between the results. Compared at an adequate grid density of 8000 points/Å^3, the computation was 4-5000 times faster, while the average for the relative differences was only 0.06% for volume and 0.08% for surface area (using the smaller radii set, where the present algorithm gives analytically exact results in many cases). With the other radii set more and bigger atomic overlaps are present making computations slightly (2-3 times) longer and results less accurate. Still, all differences are less than 1.2%, with an average of only 0.18% for volume and 0.17% for surface. Most of these differences occur in highly strained structures with multiple rings (e.g., norbornene). There was no difference in the program's behavior for differently optimized structures; any reasonable three-dimensional molecular structure could be used. The algorithm gives a sufficiently accurate and fast size estimate, and it could allow a substantial acceleration of programs using molecular size for structure-activity relationships (e.g., BLOGP, Bodor et al., 1989), molecular association, receptor binding, or other studies.

Comparing the computed volumes with readily available experimental size estimates of molecules (CRC Handbook, 1991-1992; Merck Index, 1996), such as V_m – the molar volume, or V_c – the proper volume in collision theory obtained (Partington, 1949) as $b/4$ where b is the co-volume (the b parameter in the van der Waals equation of state for a gas), both scaled for one molecule dividing them with Avogadro's number, one gets reasonable correlations that prove a

meaningful size estimate. The computed intrinsic volume is, as expected, bigger than the collision volume for a van der Waals gas, $V_c = 0.20 + 0.62 v_{Bondi}$ (with r^2 0.967 for 45 molecules), and smaller than the scaled molar volume, since even the closest packing of spheres leaves 26% of volume free, $V_m = 6.59 + 1.68 v_{Bondi}$ (with r^2 0.953 for 80 molecules). Meanwhile, the 10% reduction in radii, which results in volumes denoted v, almost does not affect these correlations (e.g., r^2 between V_m and v_{Bondi} is 0.953, while between V_m and v is 0.952). In good agreement with Leahy's observation (Leahy, 1986), the molar volume of *n*-alkanes is increased relative to other compounds with an average of about 25 $Å^3$ (15 ml/mol). This increase may result from a lack of strong interactions in liquid alkanes as no such effect is present in the collision volume.

The described algorithm combining analytical and numerical techniques allows a sufficiently fast and accurate computation of van der Waals molecular volume and surface area even for large molecules. This makes possible reasonably fast implementations of any model relying on molecular size, including those to be described in the following chapters.

CHAPTER 3
UNIFIED MODEL FOR LIQUID PROPERTIES

As a next step, the possibilities of a model are explored that allows a unified description and estimation of solubilities, partition coefficients, enthalpies of vaporization, and boiling points for a variety of organic liquids through a free energy approach. The free energy expression used is obtained from statistical mechanical concepts introducing different molecular size dependent effects. Such an approach makes unnecessary to rely on often misused concepts of standard states and activities, and it leads to relations that can not be obtained from purely thermodynamic arguments (Ben-Naim, 1980). In addition, by assuming a simple size determined interaction potential in the liquid phase, a unified and quite good description is achieved for boiling points, enthalpies of vaporization, solubilities, and partitioning properties of a large number of organic liquids that have no hydrogen bonding or strongly polar substituents. Just one structure determined variable accounts for more than 90% of the variance in these properties, and the interaction constant derived from enthalpies of vaporization can describe not only boiling points, but partition and solubility properties as well. With the inclusion of only one additional parameter, the model could be extended to reliably predict the log octanol-water partition coefficient of a wide range of compounds, including large, multifunctionalized molecules.

Background—Statistical Mechanics

Statistical mechanic theory incorporates all information about the thermodynamic properties of a system into the partition function, which represents a sum over all possible microstates *l* of energy E_l:

$$Q = \sum_l e^{-\frac{E_l}{kT}} \qquad (3.1)$$

A simple connection to usual thermodynamic quantities can be made, for example, through the Helmholtz free energy, $A = -kT \ln Q$. Thus, the chemical potential, our main interest here, is obtained for a given component *i* as (Alberty, 1987, p.636; Ben-Naim, 1980, p.260; Ben-Naim, 1978)

$$\mu_i = \left(\frac{\partial A}{\partial N_i}\right)_{T,V,N_j} = -kT\left(\frac{\partial \ln Q}{\partial N_i}\right)_{T,V,N_j} \qquad (3.2)$$

Notation is usual: k - Boltzmann constant, T - temperature, N_i - number of *i* type molecules. To obtain a model that is of any use for liquids, a few simplifying assumptions have to be made:

- *The translational degrees of freedom of the molecules are essentially classical.* This condition should be satisfied by most liquids, except maybe those that form only at very low temperatures, like liquid hydrogen.
- *The internal degrees of freedom are essentially the same in the liquid as in the gas.* This means that we assume that molecular rotation and intramolecular vibration are unaffected by the presence of neighbor molecules in the liquid.

These two assumptions are fairly general and common (Hildebrand and Scott, 1950, p 67). They allow the factoring out of the internal degrees of freedom and the integration of the translational part. Here, we omit details of the derivation and only briefly summarize the translation-related part for a nondegenarate system of indistinguishable molecules. If Boltzmann statistics applies, the canonical ensemble partition function can be written as (Alberty, 1987, p 609; Laidler and Meiser, 1995, p 718)

$$Q_t = \frac{q_t}{N!} = \frac{V^N e^N}{\Lambda^{3N} N^N} \qquad (3.3)$$

In this equation q_t is the translational part of the molecular partition function, and the value for Λ results from integration over all possible momenta of the molecule. For a particle in a box, Λ equals $(h^2/2\pi mkT)^{1/2}$ and denotes the thermal de Broglie wavelength (m - mass of molecule, h - Planck constant).

If interactions are present, some assumptions concerning the interaction potential must also be made. A frequently used one was originally introduced by Guggenheim (1932):

- *The interaction potential may be replaced by its value averaged over all accessible configurations.* This value is denoted here W_i.

This last assumption introduces an additional NW_i term in the equation for the Helmholtz free energy

$$A_t = -kT \ln Q_t + NW_i = -NkT - NkT \ln\left(\frac{V}{\Lambda^3 N}\right) + NW_i \qquad (3.4)$$

Through equation 3.2, this leads to a chemical potential that is essentially identical with that used by Ben-Naim (1980, p 262):

$$\mu_i = kT \ln\left(\frac{N_i \Lambda_i^3}{V q_i}\right) + W_i = kT\left[\ln\left(\frac{N_i \Lambda_i^3}{V q_i}\right) - \ln\left\langle e^{-\frac{B_i}{kT}}\right\rangle\right] \qquad (3.5)$$

Here B_i is the binding energy of one molecule *i* at a fixed position to all the other molecules at a specified configuration, and < > denotes an average over all the locations of the N_i molecules enclosed in the volume *V*. Furthermore, q_i is the internal partition function of a single molecule *i*, and it is considered, as stated above, independent of the environment; therefore, it will be disregarded (considered unity) from now on.

Thus, the first term in the above equation, referred by Ben-Naim as the "liberation" free energy, represents the translational contribution (Alberty, 1987) to the chemical potential μ_i, and the second term can be regarded as the average work W_i of coupling a molecule located at some fixed position to its environment.

For perfect gases the volume occupied by molecules is negligible and there are no intermolecular forces (B_i=0, W_i=0). From these conditions all familiar expressions for perfect gases can be obtained, including

$$\mu^{gas} = kT \ln\left(\frac{N}{V}\Lambda^3\right) = kT \ln(\rho\Lambda^3); \quad \Lambda = \sqrt{\frac{h^2}{2\pi m kT}} \qquad (3.6)$$

where ρ represents particle density. If neither the size, nor the interactions are negligible, a possibility is to assume that the free volume available for translation

is the total volume less that excluded by the molecules regarded as hard spheres, $V^{free} = V - Nb$, and the interaction is simply proportional with the particle density, $W = ap = aN/V$. Introducing these in equation 3.4, the well-known van der Waals equation of state for gases (Alberty, 1987, p.16) can be obtained:

$$p = -\left(\frac{\partial A}{\partial V}\right)_{T,N} = -\frac{aN^2}{V^2} + \frac{NkT}{V - Nb} \qquad (3.7)$$

Despite an overwhelming need, no thoroughly good model exists for liquids, especially for those of chemical interest. The situation is more complex here as noncovalent intramolecular forces play a major role. Due to their complex nature, it is unlikely to describe all effects (e.g., Bodor and Huang, 1992a); however, a limited number of parameters may account for a large majority of the variance in liquid properties. For example, a study by Cramer (1980) suggests that two variables suffice to explain ~95% of the variance in several physical properties of the 114 liquids considered. As molecular size clearly plays a major role, and a good number of evidence suggests linearity between size and free energies of transfer, we will explore the following assumptions accounting for different size-related effects:

- *The binding energy and thus the coupling work of a molecule to its environment in a liquid is linearly related to molecular volume.*

This should account for most nonspecific intermolecular interactions in liquids where they are, at least to a reasonable extent, orientation averaged. Therefore, for most molecules from the attractive van der Waals forces,

including dipole-dipole, dipole-induced dipole, and dispersion forces, the dispersion forces are clearly the dominant one (Grant and Higuchi, 1990; Hildebrand and Scott, 1950, p 51; Laidler and Meiser, 1995, p 813). As they strongly depend on polarizability, and polarizability is to a good extent size related, molecular volume should be a good predictor for the interaction potential of these nonspecific forces. There are many studies proving not only good correlation between free energies of transfer (as measured, for example, by log partition coefficients or log solubilities) and molecular size, but also a quite similar slope within different classes of compounds (Amidon et al., 1975; Valvani et al., 1976; Silla et al., 1992; Ben-Naim and Marcus, 1984). Therefore, instead of focusing on pairwise interactions, we introduce an average interaction energy between a molecule and its environment that is linearly related to molecular volume v, $W_i = - w_0 - w\,v$. We expect w_0 and w to be constants that are about the same for similar liquids. Size alone cannot account for interactions such as hydrogen bondings or interactions between strongly polar substituents; therefore, in a complete treatise W should contain additional terms. However, as our purpose here is to show that the present form can account for most nonspecific effects in several liquids, we will let these additional terms to be absorbed in w_0 for now.

- *The volume of the liquid phase is additive.*

The molar volume of a pure liquid will be considered proportional with the molecular volume, $V^0 = aN_0v$. This is a simplifying assumption but represents a

reasonable estimate for most organic molecules, and it makes possible estimates starting from molecular structure alone. For our data ($n = 259$) we found $V^0 = 1.24\ v$, with $r^2 = 0.90$; strong deviants being only flat, multiring aromatics such as perylene or fluoranthrene.

In addition, for mixtures of different compounds, the molar volumes will be regarded as additive: $V = n_i V_i^0 + n_j V_j^0 = a\ (N_i v_i + N_j v_j)$. Thus, the volume of the liquid phase is determined by the number and the size of the different molecules. These assumptions introduce an additional term when molecules of unequal size are mixed, term often neglected but already considered in different theories (Sharp et al., 1991; Hildebrand and Scott, 1950, p 267).

- *The volume available for translation in a liquid can be considered as a fraction f of the total liquid volume.* With this, the translational contribution to the chemical potential (eq. 3.5) becomes $kT \ln (N\Lambda^3/fV)$ with $f < 1$.

Molecular size cannot be neglected in liquid phase. Because not all the volume is available for translation, as was the case even for a van der Waals gas (eq. 3.7), the form used by Ben-Naim represents an overestimate ($f = 1$). However, the advantages of his approach can be maintained by assuming that f is approximately the same for the considered liquids. Molecules are not considered confined to cages of volume V/N at various lattice sites as in cell models (Hildebrand and Scott, 1950, p 69); they have access to the entire free volume of the liquid (Laidler and Meiser, 1995, p 816).

No temperature dependence is included in this volume or in the interaction potential, though certainly both are temperature dependent to some extent. This is again a simplification that later might affect some derived properties, even if polarizability, for example, is nearly temperature independent (Grant and Higuchi, 1990, p 67).

These assumptions give for a pure liquid:

$$\mu^{liq} = kT \ln\left(\frac{N\Lambda^3}{fV}\right) - wv - w_0 = kT \ln\left(\frac{\rho}{f}\Lambda^3\right) - wv - w_0 \qquad (3.8)$$

For a liquid mixture of *i* and *j*, the total Gibbs free energy is $G = N_i\mu_i + N_j\mu_j$. Since V is a function of N_i, the chemical potential for component *i* (considered here as the solute) obtained after performing the partial derivative is

$$\mu_i^{sol,j} = \left(\frac{\partial G}{\partial N_i}\right)_{p,T,N_j} = kT \ln\left(\frac{N_i\Lambda^3}{fV}\right) - w_{ij}\, v_i - w_0 + kT\left(1 - \frac{N_i v_i + N_j v_i}{N_i v_i + N_j v_j}\right) \qquad (3.9)$$

The term w_{ij} accounts for the interaction of solute *i* with solvent *j*, and to get a model of any use it should be about the same for similar organic solvents. As water is a very peculiar solvent, this value might be quite different for an aqueous solvent, if this description applies at all. For sufficiently dilute solutions, all *i* molecules are surrounded by *j* molecules only. No *i-i* interactions have to be considered; this justifies the use of only a w_{ij} term. Meanwhile, $N_j >> N_i$, thus

$$\mu_i^{sol,j} = kT \ln\left(\frac{\rho_i}{f}\Lambda_i^3\right) - \left(w_{ij} + \frac{kT}{v_j}\right)v_i - w_0 + kT \qquad (3.10)$$

Compared to eq. 3.8, different molecular sizes and volume additivity introduce here an additional $kT\,(1 - v/v_j)$ term. As mentioned earlier, such terms have been obtained and analyzed by different authors, and it is probably of interest to note that it is also obtained in the Flory-Huggins theory of polymer mixing (Sharp et al., 1991).

The unique and characteristic assumption of this model is that molecules in a liquid are considered moving freely in a fraction of the total volume that is not excluded by their own size under an average attractive potential of the surrounding molecules, a potential that can be described by molecular volume through a simple, linear relationship as long as no strongly specific interactions are present. To test the validity of this model, expressions for measurable properties will be derived using the chemical potentials 3.6, 3.8, 3.10, and these expressions will be compared with experimental data. The present approach relies on clearly stated assumptions regarding molecular interactions, avoids the need for often hypothetical standard states or related activities, leads to relations that cannot be obtained from purely thermodynamic arguments, and gives a unified view of all these processes as related to the equilibrium distribution of molecules between two different phases. As the model is based on computed molecular size, it makes possible reasonable property estimates from 3D molecular structures.

If a molecule partitions between two phases, equilibrium is reached when its chemical potential is the same in both phases. Equations obtained from such

conditions link the equilibrium concentrations to different intermolecular interactions in the two phases. As solubility and partitioning data used here are mainly at room temperature (T_0), we will introduce $\omega = w/kT_0$ for convenience.

Experimental Data

The complete database collected contains more than 400 molecules with at least one experimental data available. Enthalpies of vaporization, boiling points, densities, and molecular weights are from a recent collection (Handbook of Data on Common Organic Compounds, 1995); solubilities and Ostwald absorption coefficients are from Hine and Mookerje (1975), respectively from Abraham et al. (1990); partition coefficients for hexadecane-water systems are from Abraham et al. (1990, 1994), for octanol-water systems are those collected earlier mainly from the compilation of Hansch and Leo (Hansch and Leo, 1979). Most correlations were performed on a set of 115 molecules having no strongly polar or hydrogen-bonding substituents including haloalkanes, haloaromatics, and unsubstituted aromatics. A separate set of 18 alkanes and cycloalkanes was also used.

Molecular volume was computed with the previously described algorithm and the following atomic radii set: H 1.08, C 1.53, N 1.40, O 1.36, F 1.29, Cl 1.60, Br 1.83, I 2.05, S 1.70, P 1.75, Si 2.10 (all values in Å). These are about 10% smaller values than those used after Bondi (1964), but such a radii set was also tested. All conclusions were quite similar as obtained volumes correlate well enough to affect correlation coefficients with experimental data at most in the

third significant figure. However, for haloaromatics, slightly better correlations for enthalpies of vaporization, boiling points, and solubility properties were obtained after a small increase in the atomic radii of the halogen atoms attached to aromatic rings. As this was not true for molar volume, it might reflect a change in the interaction ability of these halogen atoms. A clear distinction would require data for polyhalogenated compounds, but enough data for this class are available only for octanol-water partition. As here a similar, probably even more significant trend is present, we simply used the radii set introduced in the log $P_{o/w}$ study to account for such effects (arom. F 1.25, Cl 1.80, Br 2.00, I 2.15). The van der Waals radii of these larger atoms are known to be more environment and direction sensitive (Nyburg and Faerman, 1985).

Enthalpy of Vaporization

When vaporization occurs at the boiling point, the vapor pressure equals the atmospheric pressure p_0; thus $\mu^{gas} = kT_b \ln \rho = kT_b \ln (p_0/kT_b)$, and using eq. 3.6 and 3.8 we obtain for the free energy of transfer per molecule

$$\Delta\mu = \mu^{gas} - \mu^{liq} = kT \ln\left(\frac{p_0 f}{kT \rho^{liq}}\right) + w_0 + w_{ii} v_i \qquad (3.11)$$

From here, using the Gibbs-Helmholtz equation, we obtain a molar enthalpy of vaporization that is linearly related to molecular volume:

$$\Delta H^0 = N_0 \left(\frac{\partial \left(\frac{\Delta \mu}{T} \right)}{\partial \left(\frac{1}{T} \right)} \right)_p = N_0 (kT_b + w_0 + w_{ii} v_i) = RT_0 \left(\frac{T_b}{T_0} + \omega_0 + \omega_{ii} v_i \right) \qquad (3.12)$$

Indeed, there is a very good correlation between enthalpy of vaporization and molecular volume, even if no temperature dependence for the binding energy was included in the model. Certainly, T_b also depends on size, but the available experimental data can already be used to obtain a first estimate for ω. For the compounds having no strongly polar or hydrogen-bonding substituents:

$$\Delta H^0 = RT_0 (T_b/T_0 + 5.302 + 0.084\ v) \qquad r^2 = 0.870,\ n = 34$$

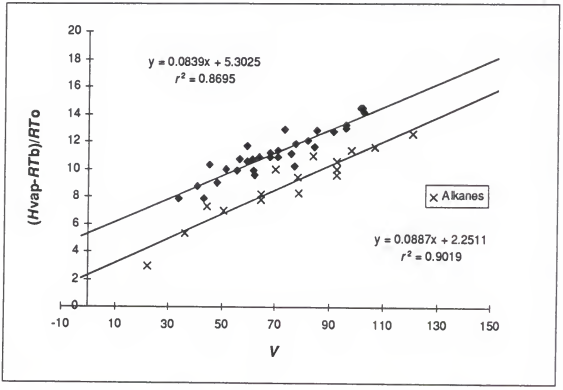


Figure 3.1 Molar enthalpy of vaporization scaled according to eq. 3.12 as a function of molecular volume (in Å^3) to obtain estimates for ω. The continuous lines are trendlines.

Compared with eq. 3.12, this gives ω_0=5.3, ω_{ii}=0.084. As it will turn out, these values can be used practically unaltered to describe all following properties (boiling point, solubility, vapor-liquid and liquid-liquid partition). Values for alkanes and slightly deviating cycloalkanes are on a line with a similar slope but a clearly different intercept

$$\Delta H^0 = RT_0 (T_b/T_0 + 2.250 + 0.089\ v) \qquad r^2 = 0.902,\ n = 16$$

suggesting $\omega_0 = 2.3$, $\omega_{ii} = 0.089$ for this group of compounds (Figure 3.1).

Boiling Point

From the condition of equilibrium that $\Delta\mu$ as expressed in eq. 3.11 is zero at the boiling point ($T = T_b$), using $\rho^{liq} = N_0/V^0_{liq}$ for the number density, we obtain after some rearranging:

$$kT_b \ln\left(\frac{k\,T_b\,N_0}{p_0\,f\,V_i^0}\right) = w_0 + w_{ii}\,v_i \qquad (3.13)$$

Even with the simple assumption that $V^0 = aN_0v$, this equation gives a complex, nonlinear dependence between boiling point (T_b) and molecular volume (v). Nevertheless, as a reasonable first estimate, T_b/V^0 can be considered constant, especially that only its logarithm appears in the equation. We have thus:

$$T_b = \frac{w_0 + w_{ii}\,v_i}{k \ln\left(\frac{RT_b}{p_0\,f\,V_i^0}\right)} = T_0 \frac{\omega_0 + \omega_{ii}\,v_i}{\ln\left(\frac{RT_b}{p_0\,f\,V_i^0}\right)} \qquad (3.14)$$

Repeating this after first adding kT_b to both sides in eq. 3.13, and using eq. 3.12 to introduce the enthalpy of vaporization, we get:

$$T_b = \frac{kT_b + w_0 + w_{ii}v_i}{k\left(1 + \ln\frac{RT_b}{p_0 f V_i^0}\right)} = \frac{\Delta H^0}{R\left(1 + \ln\frac{RT_b}{p_0 f V_i^0}\right)} \qquad (3.15)$$

The denominator is, of course, the molar entropy of vaporization at the standard boiling point, and it also can be obtained directly from $\Delta\mu$ (eq. 3.11)

$$\Delta S^0 = -N_0\left(\frac{\partial(\Delta\mu)}{\partial T}\right)_{p,N} = R\left(1 + \ln\frac{RT_b}{p_0 f V_i^0}\right) \qquad (3.16)$$

According to Trouton's rule, for most liquids this is the same constant and has a value around 88 J/K mol (Alberty, 1987, p 189). For our data its average is 87 J/K mol while the average for T_b/V^0 is 5 (±0.9) K/ml. These values give ln (RT_b/p_0V^0) = 6 and thus an estimate for f around 3% (ln f = -3.5).

It would be hard to predict exactly what portion of the total volume in a liquid can be considered as available for *free* translation, as even the closest packing of spheres leaves about 26% of the volume unfilled (Levine, 1988, p 871). There are different ways to estimate free volume within cell models where molecules are constrained to a cage; values obtained are somewhere in the 0.1-10% range (Hildebrand and Scott, 1950, p 76). The speed of sound in liquids can also be used for such an estimate. For most liquids, this speed is about five to ten times the speed of sound in gas (Laidler and Meiser, 1995, p.815). In a rough model this is so because in a liquid, where molecular size cannot be

neglected, a molecule of radius r_i instead of the whole intermolecular distance d has to travel only a d-$2r_i$ distance before hitting its nearest neighbor. The ratio of the two speeds is $(d\text{-}2r_i)/d$; therefore, the ratio of free and total volume should be somewhere around $(d\text{-}2r_i)^3/d^3$, that is around 0.1-1%. The value obtained here (3%) seems, therefore, reasonable.

Omitting the small fluoro-substituted compounds, where other more specific interactions are likely to be present, we have for the unsubstituted molecules (Figure 3.2):

$$T_b/T_0 = 0.583 + 0.0087\ v \quad r^2 = 0.878,\ n = 95$$

The equation for the trendline is not significantly changed even considering all molecules excluding alcohols, acids, and amides (well-known outliers):

$$T_b/T_0 = 0.628 + 0.0082\ v \quad r^2 = 0.775,\ n = 203$$

Alkanes are again on a separate line with

$$T_b/T_0 = 0.295 + 0.0093\ v \quad r^2 = 0.882,\ n = 17$$

Using the above derived f, we have from eq. 3.14 $\omega_0 = 5.5$, $\omega_{ii} = 0.083$, respectively $\omega_0 = 2.8$, $\omega_{ii} = 0.088$ for the alkanes – in excellent accord with values obtained from the enthalpy of vaporization ($\omega_0 = 5.3$, $\omega_{ii} = 0.084$). As more data are available here, but the expected relationship is less linear, one can consider $\omega_0 = 5.3$, $\omega_{ii} = 0.084$ (respectively $\omega_0 = 2.5$, $\omega_{ii} = 0.089$ for alkanes) as reasonable estimates.

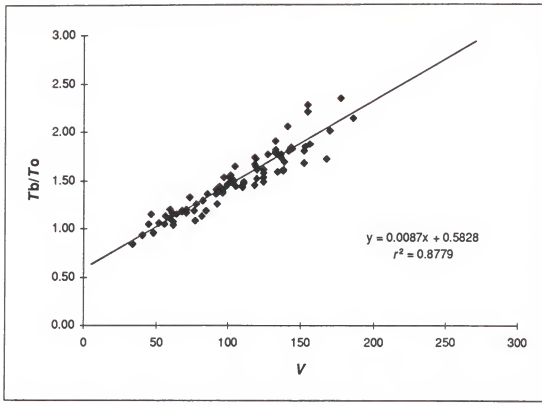


Figure 3.2 Boiling point (T_b/T_0) for 93 molecules as a function of molecular size.

Gas-Liquid Solubility and Vapor-Liquid Equilibrium

Again, at equilibrium the chemical potentials are equal

$$\mu_i^{gas} = \mu_i^{sol,1}$$

The Ostwald absorption coefficient (γ) is defined as the ratio of liquid and gas number densities (molar concentrations) at equilibrium. Using the corresponding chemical potential forms (eq. 3.6 and 3.10), we obtain for its logarithm:

$$\log\gamma_i^l = \log\left(\frac{\rho_i^{sol,l}}{\rho_i^{gas}}\right) = \frac{1}{2.303}\left[(-1 + \omega_0 + \ln f) + (\omega_{i1} + \frac{1}{v_l})v_i\right] \qquad (3.17)$$

The equation was already converted to the more common decimal logarithm. Index *i* denotes the solute; index *1* refers to the solvent. According to this equation, solubility on the molar scale is enhanced in solvents of smaller size where v_i/v_1 is larger, if all other terms can be considered equal. Such a tendency, without quantitative correspondence, was noted for solvation of small alkanes (Abraham, 1979; Ben-Naim and Marcus, 1984). The difference of two such terms ($v_i/v_1 - v_i/v_2$) will appear later in the partition equation as well, and it is exactly the correction introduced by DeYoung and Dill (1990) in describing benzene partition between water and different alkane solvents.

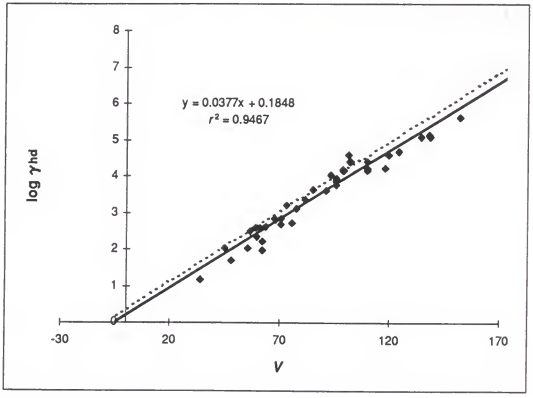


Figure 3.3 Gas-hexadecane Ostwald absorption coefficient as a function of molecular volume (in $Å^3$) for molecules included in the present database. The continuous line is the trendline while the dashed line is that predicted by the unified model having coefficients obtained from enthalpies of vaporization.

The Ostwald coefficient data in hexadecane measured by Abraham et al. (1990) give for the unsubstituted molecules

$$\log \gamma_{hd} = 0.185 + 0.0377\ v \qquad r^2=0.947,\ n=43$$

in very good agreement (Figure 3.3) with that predicted from the enthalpies of vaporization having an intercept of (-1+5.3-3.5)/2.303 = 0.347 and a slope of (0.084+1/232)/2.303=0.0383. For alkanes:

$$\log \gamma_{hd} = -0.847 + 0.0388\ v \qquad r^2=0.924,\ n=13$$

compared to (-1+2.5-3.5)/2.303=-0.868 and (0.089+1/280)/2.303=0.0402. The excellent agreement obtained this time for a different process (partition between hexadecane and gas) is good proof favoring the hypothesis of the present approach that a single, unified interaction constant can be used.

An additional advantage of such statistical mechanical approaches is that the equilibrium between a pure liquid and its vapor can be simply included by treating it as the solvation of a molecule by its own liquid, as already pointed out by Ben-Naim and Marcus (1984). Taking $1 = i$ and using $\rho = 1/V_i^0$ for the molar density of a pure liquid we have after a sign inversion:

$$\log \rho^{gas} - \log \frac{1}{V_i^0} = -\frac{1}{2.303}\left(\omega_0 - \ln f + \omega_{ii}\, v_i\right) \qquad (3.18)$$

For the 40 liquids (haloalkanes, haloaromatics, unsubstituted aromatics) where data was available:

$$\log \rho^{gas} - \log 1/V^0 = -1.157 - 0.0321\ v \qquad r^2=0.913,\ n=40$$

If we compared this with the predicted intercept, -(5.3-3.5)/2.303= -0.78, and slope, -0.084/2.303= -0.036, (Figure 3.4) agreement with the model is quite satisfactory. No regression was performed for the alkanes as their behavior is less regular and only limited data are available (smaller alkanes do not condense to liquids around room temperature).

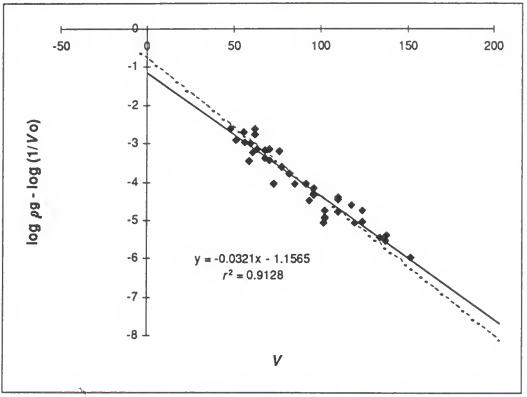


Figure 3.4 Molar gas concentration at equilibrium between a liquid and its vapor. The continuous line is the trendline while the dashed line is that predicted by the unified model having coefficients obtained from enthalpies of vaporization.

Water is certainly a special, more ordered, hydrogen-bonded liquid; a whole literature is dedicated to the description of its structure and to the description of the hydrophobic interaction. Therefore, it would be of no surprise if the interaction between a solute and an aqueous solvent could not be included

within this model. Indeed, size-dependence is quite ambiguous when describing gas-water solubility (Cramer, 1977; Abraham, 1979). The correlation between molecular size and gas-water Ostwald absorption coefficient is much worse than those obtained earlier. This indicates that other, not properly described effects are important here. Omitting the small fluoro-alkanes and the rigid aromatics (pyrene, anthracene, etc.) that are scattered away, we are left with practically no size-dependence at all:

$$\log \gamma_w = 0.5241 - 0.0006\ v \qquad r^2=0.003,\ n=43$$

With such a nonexistent correlation both the slope and the intercept depend strongly on the data included, but a very small, almost zero slope seems likely. The intercept obtained is close to that predicted by the model: (-1+5.3-3.5)/2.303 = 0.347. As the slope is close to zero, using a $(\omega_w+1/v_w)= 0$ value to describe water-solute interactions in the equations to be derived for partition between water and an organic solvent, respectively for water solubility might be a reasonable idea. The quite good fits that will be obtained suggest that – at least for this purpose – water can be included within the model using only a different interaction constant, and that the above correlation is so bad mainly because ω_w and $1/v_w$ compensate each other giving a zero slope by annihilating the volume dependence. Alkanes (excluding cycloalkanes) give again a significantly different intercept and this time a more negative slope:

$$\log \gamma_w = -1.100 - 0.0087\ v \qquad r^2=0.801,\ n=12$$

Solvation of longer chains is less favorable probably due to restricted flexibility in water. As bigger alkanes included here have longer chains, this effect might make the above slope more negative and the difference between the slope obtained for alkanes and that obtained for the other compounds larger than was for the other properties. The negative intercept obtained agrees with the model: (-1+2.5-3.5)/2.303 = -0.868.

Partition Between Two Solvents

If the two solvents are immiscible or solute partitioning is performed without allowing solvent mixing, equilibrium is reached when

$$\mu_i^{sol,1} = \mu_i^{sol,2}$$

The partition coefficient *P* is defined as the ratio of the molar concentrations in the two different phases. This being equal with the ratio of number densities, we can use eq. 3.10 for both sides and obtain:

$$\log P = \log \frac{\rho_1}{\rho_2} = \frac{1}{2.303}\left(\omega_{i1} - \omega_{i2} + \frac{1}{v_1} - \frac{1}{v_2}\right) v_i \qquad (3.19)$$

This predicts a zero intercept for log *P* represented as a function of molecular volume. A study that will be presented in the next chapter for the octanol-water system, where large number of experimental data are available, proved not only this (even for noble gases), but it also showed a very similar slope for different classes of compounds. This observation was the starting point for the hypothesis explored further here that molecular size is a good descriptor for nonspecific interactions in organic liquids.

The present data for hexadecane-water system give:

$$\log P_{hd/w} = -0.435 + 0.038\ v \qquad r^2=0.953,\ n=61$$

The intercept is more negative than the expected zero, and the slope agrees well with that predicted by the model – even if one solvent is water: (0.084+1/232+0)/2.303=0.038.

For alkanes we also have a very good correlation:

$$\log P_{hd/w} = 0.055 + 0.047\ v \qquad r^2=0.989,\ n=16$$

This time the intercept is very close to zero, and the slope agrees well with the expected one: (0.089+1/232+0.02)/2.303=0.049. Certainly a model that adequately describes gas-water and gas-hexadecane solubility will also work well here as water-hexadecane partition coefficient can be determined as the ratio these two solubility coefficients (Abraham et al., 1990). However, the good correlations obtained here prove that despite the small r^2 obtained for gas-water partition water can be included within this description using only a different ω_w.

The log octanol-water partition coefficient (log $P_{o/w}$) also fits this description, but it will be included in a separate chapter for multiple reasons:

- it is a very important parameter, widely used in quantitative structure-activity relationships (QSAR) developed for pharmaceutical, environmental, or biochemical applications;
- large number of data are available, and reliable regressions could be performed;

• the solvents involved are water and an alcohol; therefore, different ωs and different slopes are expected, but the very good correlations that will be obtained prove again the validity of this approach;

• as water and octanol have similar hydrogen bond acceptor abilities, by introducing only one additional parameter a computerized model could be developed that predicts log $P_{o/w}$ from the 3D structure, model already tested for more than 700 compounds including many multifunctionalized, biologically and chemically important molecules, even some unusually large ones;

• hopefully, this extension will serve as an example for future predictive models, derived within this approach, that will work on a larger variety of molecules for other properties as well.

Solubility (Liquid-Liquid)

A major advantage of the present approach is that the problem of solubility can be treated as partitioning between a solvent and the solute itself – as this is what happens at the molecular level. As long as the solubility is not very high, we can simply use eq. 3.19 with $2 = i$, thus

$$\log \rho_1 - \log \frac{1}{V_i^0} = \frac{1}{2.303}[-1 + (\omega_{i1} - \omega_{ii} + \frac{1}{v_i})v_i] \qquad (3.20)$$

where ρ refers again to molar concentrations, and where $\rho = 1/V^0$ was used for a pure liquid. This means that to correctly relate solubility and solute-environment interactions, connected here to molecular size, we need to consider the molar (or number) density of the pure solute, $\rho_i = n_i/V = 1/V_i^0$; a relation that may not be

obtained from purely thermodynamic arguments (Ben-Naim, 1980, p 14). When looking for connections between solubility and partition, purely thermodynamic approaches have to rely on additional assumptions regarding the activity in one or the other solvent (Yalkowsky et al., 1988; Grant and Higuchi, 1990, p.59).

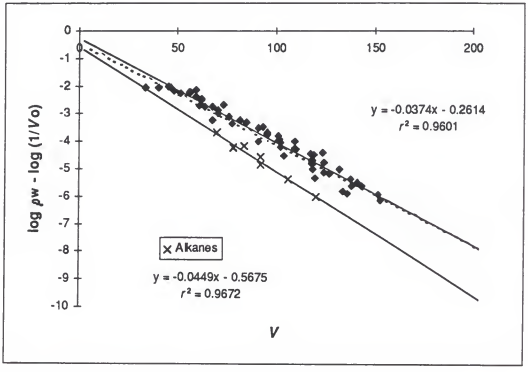


Figure 3.5 Water solubility scaled according to eq. 3.20 as a function of molecular volume (in Å^3). The dashed line is that predicted by the unified model.

Even water solubility data gives good agreement with this equation (Figure 3.5) as indicated by a comparison with values from the unified model: -1/2.303=-0.434 for intercept and (-0.084+0)/2.303=-0.0365 for slope

$$\log \rho_w - \log 1/V^0 = -0.261 - 0.0374\ v \qquad r^2=0.960,\ n=59$$

The model clearly does not work for smaller alkanes, but for those bigger than butane the agreement is good again:

$$\log \rho_w - \log 1/V^0 = -0.568 - 0.0449\ v \qquad r^2=0.967,\ n=8$$

The slope is close to that predicted: (-0.089-0.02)/2.303=-0.047.

A consistent model relying on free energy expressions obtained from statistical mechanics related molecular assumptions was presented. By incorporating two novel assumptions regarding molecular interaction and motion in liquid phase, it gives a unified treatment for various processes in liquids where no strongly specific interactions are present. An important achievement is that the molecular volume based interaction coefficient (ω) obtained from the enthalpy of vaporization can be used in the subsequently derived equations to adequately describe solubility and partition properties for this group of molecules. As these equations rely only on computable molecular volume and account for 90-95% of the variance of the corresponding properties, they might even prove useful in estimating such properties. This approach could serve as a starting point to predict properties for compounds where more specific interactions are also present. The following chapter will describe a fully computerized model developed for log P_{ow} that works even for large, multifunctionalized molecules and serves as a possible example for future developments.

CHAPTER 4
OCTANOL-WATER PARTITION

Background

The partition coefficient, defined for dilute solutions as the molar concentration ratio ($P = c_o/c_w$) of a single species between two phases at equilibrium, is an important physicochemical parameter since its logarithm, usually for the *n*-octanol/water system, is widely used as a measure of hydrophobicity/lipophilicity. Today, log $P_{o/w}$ probably represents the most informative and successful physicochemical property used in medicinal chemistry (Testa et al., 1996); it has been extensively employed to explain drug-receptor and drug-biological membrane interactions, and it has been a crucial parameter in many quantitative structure-activity relationships (QSAR) developed for pharmaceutical, environmental, biochemical, or drug design applications (Lipinski et al., 1997). Ever since Overton (1897, 1901) and Meyer (1899) discovered a century ago that narcotic potency in a set of congeners tends to increase as the oil/water partition coefficient increases, there has been a more or less continuous interest in defining lipophilicity and its role in biological activity. Three decades later, Collander showed that nonelectrolyte permeability in membranes of the plant cell *Chara ceratophylla* correlates well with the olive oil/water partition coefficient (Collander and Bärlund, 1933;

Collander, 1937), an observation that became the basis of the lipoid theory of the cell membrane. Yet, this field started to expand only in the 60's following the work of Hansch (Hansch et al. 1962; Fujita et al., 1964), and the first review devoted entirely to partition coefficients was published only in 1971 (Leo et al., 1971). Since its first use by Collander (1951), and especially during the last three decades, log $P_{o/w}$ has became the standard lipophilicity parameter used in such QSAR and other relationships. Logarithms are employed because of the wide range to be covered, often close to 8-10 orders of magnitude, and because of the theoretical justification of linear free-energy relationships. Selection of the octanol/water system as reference is usually rationalized as modeling partition between aqueous- and biophases (Hansch and Dunn, 1972), but the proper choice of model solvent system is now again quite open; notable attempts being the use of cyclohexane/water (Stein, 1986) or Δlog P between octanol/water and cyclohexane/water (Young et al., 1988; Abraham et al., 1994b). The main reasons favoring the octanol/water system are as follows:

- Large amounts of experimental data are already available (Hansch and Leo, 1979; Sangster, 1989; Hansch et al., 1995).

- Most compounds of interest have log $P_{o/w}$ values that are in a range where experimental determinations using standard (e.g., shake-flask) methods are relatively simple (Sangster, 1989; Taylor, 1990). This is a consequence of the fact that octanol still retains some of the hydrogen-bonding ability of water; in very nonpolar solvents, like hexadecane, structural variations tend to cause

large variations and solute association can easily become a problem. However, even with octanol, measurements can be quite time consuming, and automated techniques cannot do much when solubility problems have to be overcome; one more reason why reliable computerized predictions are needed.

• Several approaches for log $P_{o/w}$ estimation have already been developed. In most cases, quite reliable quantitative predictions are possible; this is a clear advantage compared to, e.g., reverse phase HPLC retention factors where so far there is no general method that allows estimation ahead of synthesis (Taylor, 1990). Some of these predictive methods are as follows:

* atomic or group contribution methods
    - ◊ Rekker's pioneering fragmental *f* method developed using a "reductionist approach", where, after defining a set of fragments, their fragmental contributions were obtained by linear regression (Nys and Rekker, 1974, 1974; Rekker, 1977; Rekker and de Kort, 1979; Rekker and Mannhold, 1992)
    - ◊ the widely used similar method of Hansch and Leo (CLOGP – Leo et al., 1975; Hansch and Leo, 1979) developed using a "constructionist approach" that starts with the fragment value of H and each fragment's contribution is determined based on one corresponding well-determined experimental value. This method has been adapted for computational use (Chou and Jurs, 1979) and is continuously

"improved" by introduction of new correction factors (Leo, 1983, 1990; Calvino et al., 1992).

- the atomic approach of Broto et al. (1984) using 222 descriptors, followed by that of Ghose and Crippen with 110 descriptors (ALOGP – Ghose and Crippen, 1986; Ghose et al., 1988; Viswanadhan et al., 1989), and recently by that of Wang et al. using 80 descriptors (XLOGP – Wang et al., 1997). They all include additional correction factors as well.
- the group contribution method of Suzuki and Kudo (1990) or that implemented using SMILES codes as input (Smilogp – Convard et al., 1994)
- the atom/fragment contribution method (AFC, LOGKOW) of Meylan and Howard (1995)
- the log *P* evaluation method of Klopman et al. using their computer automated structure evaluation (CASE) approach (KLOGP – Klopman and Wang, 1991; Klopman et al., 1994),
- the computerized ACD/LogP approach that uses over 500 different "fragmental contributions" and over 2000 different "intermolecular interaction contributions" in its present version (Advanced Chemistry Development Inc., 141 Adelaide St. West, Suite1501, Toronto, Ontario, Canada)

* models using estimates of molecular size together with other structural parameters (Moriguchi et al., 1976; Iwase et al., 1985), and a newer "simple" method using atom/fragment-based contributions (Moriguchi et al., 1992)
* a model based on algorithmically derived graph theoretic variables (Niemi et al., 1992);
* the solvatochromic model developed by Taft, Kamlet, and coworkers (Kamlet et al., 1983, 1988; Taft et al., 1985b, 1996; Leahy, 1986) that uses a more rigorous physicochemical basis to describe such properties, but the parameters are either derived from solvent effects or calculated by a rather complex set of (apparently empirical) rules with values "settled down" after a series of successive approximations;
* an interpretative model derived from mobile order thermodynamics that needs, however, estimates of molar volumes, Hildebrand solubility parameters, etc. (Sarraf, 1997)
* models based on molecular orbital or semiempirical quantum-chemical calculations (Rogers and Cammarata, 1969a-b; Klopman and Iroff, 1981; BLOGP – Bodor et al., 1989, 1994; Bodor and Huang, 1992b)
* a recent model based on *ab initio* quantum chemical calculations at HF/6-31G* level (Haeberlein and Brinck, 1997) that uses three parameters that could be interpreted as representing a cavity term, a dipolarity/polarizability term, and a hydrogen bonding term

* computationally intensive methods, such as the solvent-dependent conformational analysis procedure (SCAP) of Hopfinger and Battershell (1976) or the attempts for simple organic compounds using Monte Carlo simulations (Jorgensen et al., 1990; Dunn et al., 1991; Dunn and Nagy, 1992) that are, however, difficult to parametrize for this partitioning system;
* some recent approaches attempting to introduce three-dimensionality by combining atomic contributions and conformational effects. A number of approaches are using the atomic contributions of Ghose and Crippen: the solvent-accessible surface area scaled atomistic method of Masuda et al. (1997), the molecular hydrophobicity mapping approach (Heiden et al., 1993) and its extension into a free energy surface density concept (Pixner et al., 1994). Hydrophobic atom constants derived from Leo's fragment constants are also used (HINT – Kellogg et al., 1991).
* "molecular lipophilicity potential" approaches, a concept introduced by Audry et al. (1986) using a distance dependent form with a long range behavior resembling that of a Coulomb potential. Fauchère et al. (1988) used a modified, exponential distance dependence, while Furet et al. (1988) replaced the fragmental values used with Ghose-Crippen atomic contributions, and later several other applications have been published (Gaillard et al., 1994; Testa et al., 1996). Despite the complete lack of physical basis for the introduced distance dependencies or for the use of the atomic log *P* contributions as "partial charges," not to mention the problems

behind the breakdown of a molecular property like log *P* into atomic or fragment contributions, they had some limited successes.

* attempts to introduce conformation-dependency with a program in which a series of experimental free energies are expressed as functions of the solvent-accessible surface of the fragments (HYDRO – Richards et al., 1991, 1992) or with a method that relies on atomic parametrization and conformation-dependent descriptors such as surface, atomic charges, dipole moment, lone-pair index (Kantola et al., 1991; Alkorta and Villar, 1992);
* finally, the present, remarkably simple, fully computerized, molecular size-based approach (Bodor and Buchwald, 1997a; Buchwald and Bodor, 1997).

• Starting with the seminal work of Hansch (e.g., Hansch et al., 1962, 1987), many successful correlations have been established especially when CNS effects are considered. For example, Franks and Lieb (1978) found that for anesthetic potency plotted against solvent/water partition coefficient octanol gave an excellent correlation for all compounds and all animals over an extremely wide range. Correlations were much better than for oil/water, hexadecane/water, or benzene/water systems – even if, for example, benzene has a solubility parameter very close to that of octanol ($\delta \approx 9$ cal$^{1/2}$ cm$^{-3/2}$). As general anesthetics were found to have no effect on lipid bilayer structures when studied using X-ray and neutron diffraction, the authors even concluded that this suggests that a single site of action is sufficient to account for this data and this

site has solubility characteristics similar to those of octanol. The logarithm of cerebrovascular permeability, estimated from the decrease of arterial plasma concentration (corrected for plasma protein binding), also correlates quite well with log $P_{o/w}$ over a wide range of about 8 log units as Figure 4.1 indicates based on partition (Hansch et al., 1995) and permeability data (Rapoport et al., 1979; Levin, 1980; Pardridge et al., 1990; Begley, 1992; Greig, 1992) collected from the literature. Some compounds, such as cyclosporin A, vinblastine, adriamycin, or epipodophylotoxin were exceptions, but it is now well established that practically all of them are substrates of P-glycoprotein, a multidrug transporter that is actively excluding them from the cerebral endothelial cells (Tsuji et al., 1993; Begley, 1996; Tsuji and Tamai, 1997; Van Asperen et al., 1997). Countless other examples where log $P_{o/w}$ proved a useful descriptor are available in the literature (Hansch and Dunn, 1972; Hansch and Clayton, 1973; Seydel and Schaper, 1982). Since several biological effects of a solute, including its cell membrane permeability, undergo a rise and fall as its partition coefficient between an organic solvent and water increases, such relationships have been frequently considered as being approximately parabolic ones (Hansch and Clayton, 1973 - review; Hansch et al., 1968; Flynn and Yalkowski, 1972; Timmermans et al., 1977; Biagi et al., 1980; Schoenwald and Huang; 1983). Mathematically such a parabolic dependence between some biological activity and log partition coefficient

$$1/c = a_1 (\log P)^2 + a_2 \log P + a_3 \qquad (4.1)$$

can be obtained assuming that the probability for a drug molecule to reach the active site in the allotted time is normally distributed with respect to log *P* around an ideal log P_0 value (Hansch and Clayton, 1973). If this ideal log P_0 is sufficiently far from the studied range, a linear relationship should be observed; linearity that is indeed frequently observed in many studies (Hansch and Dunn, 1972 - review).

As already enumerated, many methods for estimating log *P* are described in the literature, but most of them are "fragment constant" methods, in which a structure is divided into previously defined fragments and the corresponding contributions are summed together to yield the final log *P* estimate. It is rather interesting that such a complex molecular property is, even today, calculated mainly on the basis of simple, empirically derived fragment values, ignoring completely the processes involved in solvation and the molecular properties affecting them. As these methods lack a sound scientific basis, they need to introduce hundreds of fragments and additional correction factors. As an extreme example ALOGP distinguishes between 10(!) different hydrogen atom contributions with values from -0.326 to +0.442 (Viswanadhan et al, 1989, 1993).

We present here a model derived from that described earlier. It achieves a remarkable reduction in the number of parameters while it performs no worse than most other more complex methods. Its success, which was achieved using a 3D parameter (molecular volume) and a structure determined, quantified parameter, proves that the "two-dimensional" approaches were acceptable only

when the computation of 3D molecular structure was computationally too demanding. The inclusion of a three dimensional parameter seems to eliminate the need for many different, empirically derived fragmental constants.

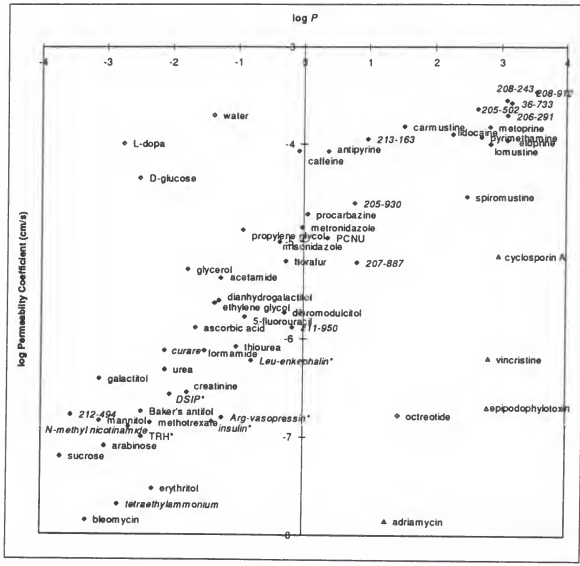


Figure 4.1 Correlation between *in vivo* permeability of rat brain capillaries and log octanol/water partition coefficient (log *P*), a measure of molecular lipophilicity. For compounds in italic, the log distribution coefficient measured at physiological pH was used (log *D*). Practically all strong deviants below the line (denoted with triangles) have been shown recently to be substrates for P-glycoprotein, a multidrug transporter that actively removes them from the brain. The graphic also suggests that octreotide, the only exception, is a likely P-glycoprotein substrate as well.

Experimental Data

The experimental data used is mainly from the compilation of Hansch and Leo (1979) with some additions from the paper by Sangster (1989) and other references (Tomida et al., 1978; Karichkoff et al., 1979; Tewari et al., 1982; Miller et al., 1984; Shiu and Mackay, 1986; Hansch et al., 1987; Leo, 1990; Debnath et al., 1991; Kossoy et al., 1992; Morimoto et al., 1992; El Tayar et al., 1993; Leo, 1993) using the average value when more acceptable measurements were available. The data set was selected so as to contain as many different functional groups as possible and to include many molecules of pharmaceutical interest, even compounds with hypervalent sulfur. Regressions were performed with a standard spreadsheet program (Microsoft Excel 5.0).

The van der Waals volume and surface area used was computed with the following radii set: H 1.08, C 1.53, N 1.40, O 1.36, F 1.29, Cl 1.60, Br 1.83, I 2.05, S 1.70, P 1.75, Si 2.10 (all values in Å). A better fit in log *P* for halogenated aromatic compounds was obtained after a slight increase in the atomic radii of larger halogen atoms attached to aromatic rings (arom. F 1.25, Cl 1.80, Br 2.00, I 2.15). The van der Waals radii of these larger atoms are known to be more environment and direction sensitive (Nyburg and Faerman, 1985); the need for a change here might also reflect a change in their interaction ability with the surrounding solvent molecules.

Model—Results and Discussion

Model for Log $P_{o/w}$

As results from the previous chapter, one can expect a good linear relationship between log *P* and molecular size; using volume or surface area makes essentially no difference for not very large organic molecules, as discussed earlier. Many good correlations (Herman, 1972; Harris et al., 1973; Pearlman, 1986; Silla et al., 1992; Leo, 1993) within one or another class of compounds, i.e., alkanes (Leo, 1993), alkyl- or halo-substituted aromatics (Pearlman, 1986), oxygen-containing organic compounds (Amidon et al., 1975; Valvani et al., 1976; Silla et al., 1992), etc. prove the correctness of these assumptions. Nevertheless, none of them could be simply extended for a larger variety of compounds, despite some indications (Grant and Higuchi, 1990) that even if many processes are involved in solvation (cavity formation, solvent reorganization, dipole-dipole and dipole-induced dipole interactions, hydrogen bonding) two parameters may account for a large part of the variance in log $P_{o/w}$. This is supported by the principal component analyses of Dunn et al. (1986), by Cramer's study on liquid properties (Cramer, 1980), by the finding that only two of the six so-called TLSER parameters are significant when correlating with log $P_{o/w}$ (Famini et al., 1992), respectively by the studies related to Taft and Kamlet's solvatochromic model (Taft et al., 1985; Kamlet et al., 1987, 1988), by El Tayar et al. (1992), and to some extent by Yalkowsky et al. (1988), but no model

working on a large variety of solutes and relying on such a limited number of parameters had been yet described.

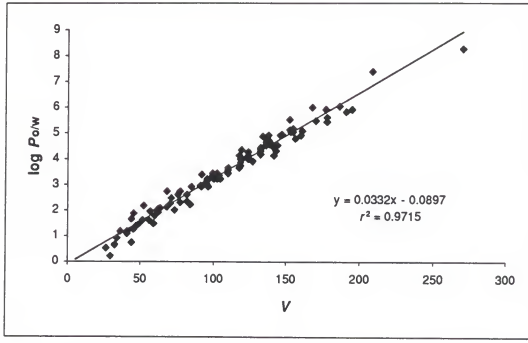


Figure 4.2 Octanol-water partition coefficient as a function of molecular volume for 115 compounds having no strongly polar or hydrogen-bonding substituents.

In the present database, for molecules containing no strongly polar or hydrogen-bonding substituents, despite neglecting solvent-dependent conformation (Hopfinger and Battershell, 1976) and other effects, we have a very good fit, a zero intercept, and a slightly smaller slope than for the hexadecane-water system (Figure 4.2):

$$\log P_{o/w} = -0.090 + 0.0332v \qquad r^2 = 0.971,\ n = 115$$

and for alkanes:

$$\log P_{o/w} = 0.350 + 0.0376v \qquad r^2 = 0.987,\ n = 19$$

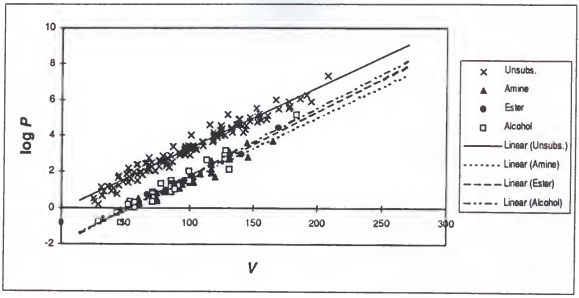


Figure 4.3 Log $P_{o/w}$ as a function of molecular volume for different classes of monosubstituted compounds.

A model for a large variety of molecules can be developed by accepting $(\omega_w+1/V_w)= 0$ for water and $(\omega_o+1/V_o)= 0.033\times2.303$ for octanol-solute, values slightly smaller than used before. In agreement with eq. 3.19, the intercept obtained, -0.090 ± 0.063, does not differ significantly from zero suggesting that as solute size decreases log $P_{o/w}$ for nonfunctionalized molecules goes toward zero. Octanol-water partition data on noble gases (Abraham et al., 1994) confirms this, as both McGowan's characteristic volume (-0.016 ± 0.039) and that computed with the radii from Bondi's 1964 paper (0.059 ± 0.126) give practically zero intercepts. It should be noted here that disagreement as to proper units and standard states is quite common in the area of solute partitioning. A zero intercept agrees well with studies (Ben-Naim, 1978; Sharp et al., 1991), suggesting that the adequate concentration scale to use in free

energy–partition relations is the molarity, comparing equal volumes as is the case here, and not the mole fraction scale that would suggest different entropies of mixing due to different solvent mole numbers.

Searching for adequate parameters to extend the model for other molecules, after noticing that the slope (figure 4.3) was quite similar for the different classes of compounds (except the alkanes), we computed the residuals in log *P* and grouped them for different classes of monosubstituted compounds. We computed thus log *P* - 0.0332*v*, where 0.0332 was determined by the linear regression mentioned earlier. Surprisingly, the obtained residuals are quite similar within each class (i.e., alcohols, ketones, amines, etc.), and they averages – as shown in figure 4.4 with the error bars indicating the standard deviations and the data labels the number of respective compounds within each group – are practically identical (~ -1.50) for almost all oxygen- or nitrogen-containing aliphatic monofunctionalized molecules. The value for amides is twice those of the others. The average for amines is slightly more negative, but the only significant exception is the aldehyde group, no doubt due to formation of hydrate, $RCH(OH)_2$ (Abraham et al., 1994). Such a good agreement is certainly due to a balancing of interactions in the two phases, but on such a variety of structures and a relatively large dataset it must be more than pure coincidence. We believe that it suggests some common mechanisms and represents a good argument against the many existent multifragment methods. For aromatic compounds, where electron withdrawing and electron donating influences play a

more important role, the pattern is more varied, but the existence of some regularity around multiples of ~-0.75 (half the above value) can still be recognized.

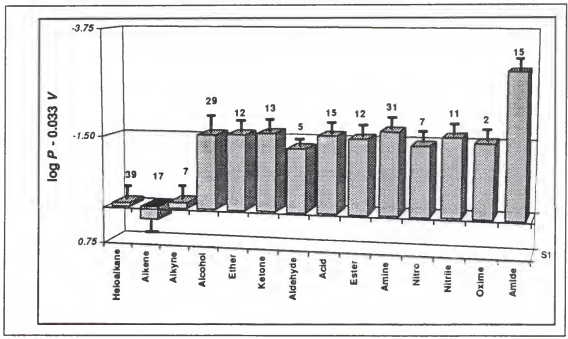


Figure 4.4 Residuals in log $P_{o/w}$ for different classes of monofunctionalized molecules after subtracting the size dependent part as suggested by the present model.

Therefore, we introduced a parameter ($\mathcal{H}$) that allowed the development of a surprisingly simple model connecting molecular size and log P for a wide range of compounds. The core equation used is

$$\log P = a_0 + a_1 V + a_2 \mathcal{H} \qquad (4.2)$$

where $\mathcal{H}$ is a positive integer increased in an additive manner by each functional group of the molecule. As results from the previous paragraph, essentially all polar, oxygen- and nitrogen-containing simple functions present in the solute

molecule require an increase by 2 units in π while those in an aromatic environment require just an increase by 1. Values used for most commonly encountered functions are as follows: 2 for -OH, -O-, $-NH_2$, -NH-, -N<, -CN, $-NO_2$, -CO-, -COOH, -COO-, -CH=NOH, -N=N-; 4 for -CON<, -CONH-. For those attached to an aromatic ring: 1 for Ar-OH, Ar-O-, Ar-NO_2, Ar-CN, Ar-COOH, 2 for Ar-NH_2, Ar-NH-, Ar-COO-, 3 for Ar-CONH-, 4 for Ar-SO_2NH_2. The contribution of each function is fixed, and for most molecules the assumption of a simple additivity works well. In some cases, like amides or aromatic nitrogens, these values are diminished with one or two units by electron-withdrawing substituents (e.g., $-NO_2$, $-CF_3$) attached directly or to the same aromatic ring. For example, π = 2 for aniline, but π = 1+1 and not 1+2 for *p*-nitroaniline. The rules describing the contribution of the different functional groups to π are sufficiently clear and consistent to be easily implemented in a computer program. The program in the actual stage recognizes some 40 different oxygen-, nitrogen-, or sulfur-containing functions and determines the corresponding integer contributions to π. All other modifications in the molecule, like branching or halogen substitution, are taken into consideration only via their effect on the volume. The examples in figure 4.5 illustrate how these rules are applied and prove the inherent simplicity of the method.

Although partition coefficients of alkanes are of very little practical importance, we have decided to include them in the general scheme. However, their log *P* increases faster with size than predicted. The inclusion of a simple

indicator variable (η) that is 1 for all saturated, unsubstituted hydrocarbons and 0 otherwise was not enough, but with a correction to the slope ($a_3 V\eta$) that is nonzero only for these compounds they also could be well described. This behavior might be the result of their unique character. A lack of strong interactions affects their solubility, especially that in water, and it also represents a probable cause behind the increase occurring in their molar volume (Leahy, 1986). This additional increase in slope occurring for alkanes 0.010(±0.0007) agrees very well with the constants obtained in chapter 3: [(0.089-0.084)+0.020]/2.303 = 0.0108.

The final model,

$$\log P = 0.032(\pm 0.0002)V - 0.726(\pm 0.007)\pi + 0.010(\pm 0.0007)V\eta \qquad (4.3)$$

obtained after regression on a set of 320 different molecules gives a remarkable $r = 0.9898$ and $\sigma = 0.214$. Volume gave a slightly better correlation than surface area, and the radii set used here gave a practically zero intercept; hence, by imposing a zero intercept ($a_0 = 0$), we obtained a very consistent model with only three adjustable parameters. After implementing the computer program, its predictive power was tested on a set of 438 new molecules ranging from hydrogen to prednisolone, sulfanilamide, or propicillin. The correlation between predicted and experimental log P is still quite impressive for such a simple model on such a wide range of compounds: $r = 0.975$, $\sigma = 0.365$. Results are presented in figure 4.6, with representative data in table 4.1.

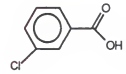
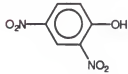
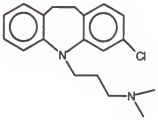
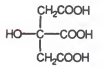
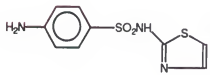
Compound formula and name		v	π	log P_{pred} = 0.032v-0.72π	log P_{exp}
CCl_2CCl_2	tetrachloroethylene	76.87	0	2.45	2.60
$(CH_3)_2CHCH_2CH_2OH$	isopentyl alcohol	85.57	2	1.28	1.29
$(C_4H_9)N(CH_3)_2$	butyl dimethyl amine	102.77	2	1.82	1.70
	3-chlorobenzoic acid	107.16	1	2.69	2.68
	2,4-dinitrophenol	109.56	1+1+1	1.32	1.53
	chlorimipramine	255.04	2+2	5.22	5.19
	citric acid	122.70	4x2	-1.88	-1.72
	sulfathiazole	156.23	2+4+1	-0.09	0.05

Figure 4.5 Illustrative sample for the presented log *P* model.

For some not completely clear reason the partition coefficient seems to be systematically overestimated for a small group of large, special molecules with connected ring structures and only a few polar substituents, such as testosterone, deoxycorticosterone, atropine, morphine, pentazocine, and analog structures. The reason behind it is not completely clear. A possibility would be that the rigidity restricts internal rotation and affects the entropy change at solvation in water for aliphatic chains, resulting in a lower lipophilicity than calculated, as noted, for example, by Osinga (1979) on the basis of Cramer's (1977) study. Many of these molecules have log *P* values that caused much agonizing in the literature; for the present model, error seems to occur when the v/π ratio becomes too large. Using a correction proportional to this ratio ($-0.025v/\pi$), quite good values are obtained for most cases. The present version will suggest this correction when v/π is larger than a threshold and a connected ring structure is present. About 20 such molecules were not included in the regression; a representative sample is included at the end of table 4.1. It is worth mentioning that this correction works even for Δ^1-tetrahydrocannabinol, a controversial molecule whose partition coefficient is strongly mispredicted by practically all methods (log P_{exp} = 3.78, while here after correction log P_{pred} = 3.66). The model still can be considered an essentially two-parameter one, as an additional correction has to be used only for a very limited number (about 5%) of quite special molecules (alkanes and those described here).

Table 4.1 Experimental and Calculated log *P* for Representative Compounds Included in the Present Study.

Compound	Formula	log *P*	*v*	*q*	π	log P_{calc}
Chloroform	$C_1H_1Cl_3$	1.97	56.58	0	0	1.80
Iodomethane	$C_1H_3I_1$	1.60	51.63	0	0	1.64
Methylamine	$C_1H_5N_1$	-0.57	32.57	0	2	-0.41
Tetrachloroethylene	C_2Cl_4	2.60	76.87	0	0	2.45
1,1,1-Trichloroethane	$C_2H_3Cl_3$	2.49	70.80	0	0	2.25
2,2,2-Trifluoroethanol	$C_2H_3F_3O_1$	0.37	53.45	0	2	0.26
Acetonitrile	$C_2H_3N_1$	-0.34	36.16	0	2	-0.29
Ethylene oxide	$C_2H_4O_1$	-0.30	37.67	0	2	-0.25
Acetic acid	$C_2H_4O_2$	-0.27	44.14	0	2	-0.04
Bromoethane	$C_2H_5Br_1$	1.61	54.74	0	0	1.74
Acetaldoxime	$C_2H_5N_1O_1$	-0.13	48.12	0	2	0.09
Nitroethane	$C_2H_5N_1O_2$	0.18	54.48	0	2	0.29
Ethanol	$C_2H_6O_1$	-0.27	43.33	0	2	-0.07
Dimethylsulfoxide	$C_2H_6O_1S_1$	-1.35	58.38	0	4	-1.03
Chloropropane	$C_3H_7Cl_1$	2.04	61.96	0	0	1.97
Trimethylamine	$C_3H_9N_1$	0.27	60.84	0	2	0.49
Furan	$C_4H_4O_1$	1.34	52.30	0	0	1.66
Crotonic acid	$C_4H_6O_2$	0.72	66.23	0	2	0.66
N,N-Dimethyl acetamide	$C_4H_9N_1O_1$	-0.77	75.70	0	4	-0.48
Butyramide	$C_4H_9N_1O_1$	-0.21	75.01	0	4	-0.50
Butane	C_4H_{10}	2.89	64.55	1	0	2.71
Isobutane	C_4H_{10}	2.76	64.59	1	0	2.72
N-Nitrosodiethylamine	$C_4H_{10}N_2O_1$	0.48	84.72	0	3	0.53
Isobutyl alcohol	$C_4H_{10}O_1$	0.75	71.52	0	2	0.83
Diethyl ether	$C_4H_{10}O_1$	0.83	71.96	0	2	0.84
Ethyl sulfide	$C_4H_{10}S_1$	1.95	81.12	0	1	1.86
3-Chloropyridine	$C_5H_4Cl_1N_1$	1.33	81.39	0	2	1.14
Cyclopentane	C_5H_{10}	3.00	70.10	1	0	2.95
3-Pentanone	$C_5H_{10}O_1$	0.99	79.79	0	2	1.09
Propyl acetate	$C_5H_{10}O_2$	1.24	86.91	0	2	1.32
Piperidine	$C_5H_{11}N_1$	0.85	80.33	0	2	1.11
1,3,5-Trichlorobenzene	$C_6H_3Cl_3$	4.02	119.62	0	0	3.81
3-Chloronitrobenzene	$C_6H_4Cl_1N_1O_2$	2.44	102.87	0	1	2.55
2-Cyanopyridine	$C_6H_4N_2$	0.50	78.00	0	3	0.31
2,4-Dinitrophenol	$C_6H_4N_2O_5$	1.53	109.56	0	3	1.32
Fluorobenzene	$C_6H_5F_1$	2.27	70.72	0	0	2.25
Nitrobenzene	$C_6H_5N_1O_2$	1.84	85.76	0	1	2.01
Benzene	C_6H_6	2.14	68.03	0	0	2.16
2-Fluoroaniline	$C_6H_6F_1N_1$	1.26	80.12	0	2	1.10
Nicotinamide	$C_6H_6N_2O_1$	-0.37	88.87	0	5	-0.78
p-Nitroaniline	$C_6H_6N_2O_2$	1.39	95.00	0	2	1.58
Phenol	$C_6H_6O_1$	1.48	74.21	0	1	1.64
Thiophenol	$C_6H_6S_1$	2.52	83.31	0	0	2.65
Aniline	$C_6H_7N_1$	0.91	77.41	0	2	1.02

Table 4.1—continued

Compound	Formula	log *P*	*v*	*y*	*z*	log P_{calc}
Benzenesulfonamide	$C_6H_7N_1O_2S_1$	0.31	98.05	0	4	0.23
Ampyzine	$C_6H_9N_3$	0.96	98.06	0	3	0.95
1,5-Hexadiene	C_6H_{10}	2.45	81.97	0	0	2.61
Cyclohexanol	$C_6H_{12}O_1$	1.23	90.94	0	2	1.45
2,3-Dimethylbutane	C_6H_{14}	3.85	92.78	1	0	3.90
Myleran	$C_6H_{14}O_6S_2$	-0.52	160.46	0	8	-0.67
3-Bromobenzoic acid	$C_7H_5Br_1O_2$	2.87	115.19	0	1	2.94
4-Chlorobenzoic acid	$C_7H_5Cl_1O_2$	2.65	107.13	0	1	2.69
Phenyl isothiocyanate	$C_7H_5N_1S_1$	3.30	98.86	0	0	3.14
Benzimidazole	$C_7H_6N_2$	1.38	87.32	0	2	1.33
Benzoic acid	$C_7H_6O_2$	1.85	89.86	0	1	2.14
Benzamide	$C_7H_7N_1O_1$	0.65	92.77	0	3	0.78
o-Hydroxybenzamide	$C_7H_7N_1O_2$	1.28	98.91	0	3	0.98
p-Hydroxybenzamide	$C_7H_7N_1O_2$	0.33	99.12	0	4	0.26
Toluene	C_7H_8	2.57	82.00	0	0	2.61
Phenylurea	$C_7H_8N_2O_1$	0.87	102.06	0	3	1.08
o-Methoxyphenol	$C_7H_8O_2$	1.33	94.97	0	2	1.58
N-Methylaniline	$C_7H_9N_1$	1.71	91.64	0	2	1.47
o-Methoxyaniline	$C_7H_9N_1O_1$	0.95	98.15	0	3	0.96
2-Heptanone	$C_7H_{14}O_1$	1.98	107.86	0	2	1.99
m-Cyanobenzamide	$C_8H_6N_2O_1$	0.52	106.67	0	4	0.50
Benzylthiocyanate	$C_8H_7N_1S_1$	1.99	107.95	0	2	1.99
Styrene	C_8H_8	3.05	91.16	0	0	2.90
Acetophenone	$C_8H_8O_1$	1.63	97.51	0	2	1.66
Phenoxyacetic acid	$C_8H_8O_3$	1.30	110.55	0	3	1.35
Acetanilide	$C_8H_9N_1O_1$	1.21	106.85	0	3	1.23
m-Xylene	C_8H_{10}	3.20	95.99	0	0	3.05
Dimethylphenylsilane	$C_8H_{12}Si_1$	3.99	123.16	0	0	3.92
Ethyl benzoate	$C_9H_{10}O_2$	2.49	118.44	0	2	2.32
2-Naphthol	$C_{10}H_8O_1$	2.81	110.60	0	1	2.80
1-Naphthylamine	$C_{10}H_9N_1$	2.24	113.78	0	2	2.17
1,2,4,5-Tetramethylbenzene	$C_{10}H_{14}$	4.00	124.28	0	0	3.95
4-Phenylpyridine	$C_{11}H_9N_1$	2.55	123.04	0	2	2.47
Sulfapyridine	$C_{11}H_{11}N_3O_2S_1$	0.00	164.29	0	7	0.17
Ethyl cinnamate	$C_{11}H_{12}O_2$	2.95	140.44	0	2	3.02
Pentobarbital	$C_{11}H_{18}N_2O_3$	1.99	176.43	0	5	2.00
Phenanthrene	$C_{14}H_{10}$	4.52	140.79	0	0	4.48
Cephalothin	$C_{16}H_{16}N_2O_6S_2$	1.01	261.17	0	10	1.08
Azidocillin	$C_{16}H_{17}N_5O_4S_1$	2.29	257.70	0	8	2.42
Benzylpenicillin	$C_{16}H_{18}N_2O_4S_1$	1.75	238.59	0	8	1.81
Propranolol	$C_{16}H_{21}N_1O_2$	3.37	211.77	0	5	3.12
Methopromazine	$C_{18}H_{22}N_2O_1S_1$	4.46	246.53	0	5	4.23
Imipramine	$C_{19}H_{24}N_2$	4.66	237.85	0	4	4.68
Prednisone	$C_{21}H_{26}O_5$	1.46	274.50	0	10	1.51
Hydrocortisone	$C_{21}H_{30}O_5$	1.61	286.00	0	10	1.87

Table 4.1—continued

Compound	Formula	log *P*	*v*	*?*	*%*	log P_{calc}
Predicted values						
Ethyleneglycol	$C_2H_6O_2$	-1.65	50.27	0	4	-1.29
Acrylonitrile	$C_3H_3N_1$	0.11	45.15	0	2	-0.01
Oxazole	$C_3H_3N_1O_1$	0.12	47.48	0	2	0.07
Allyl alcohol	$C_3H_6O_1$	0.17	52.16	0	2	0.21
2-Propanol,1-amino	$C_3H_9N_1O_1$	-0.96	67.65	0	4	-0.74
5-Fluorouracil	$C_4H_3F_1N_2O_2$	-0.88	74.09	0	5	-1.25
1,3-Butadiene	C_4H_6	1.99	53.38	0	0	1.70
Vinyl acetate	$C_4H_6O_2$	0.73	67.07	0	2	0.69
4-Chloropyridine	$C_5H_4Cl_1N_1$	1.28	81.40	0	2	1.14
3-Methyl-2-butanol	$C_5H_{12}O_1$	1.28	85.41	0	2	1.27
Pentachlorobenzene	$C_6H_1Cl_5$	5.03	152.85	0	0	4.86
4-Bromophenol	$C_6H_5Br_1O_1$	2.56	99.51	0	1	2.44
Sulfanilamide	$C_6H_8N_2O_2S_1$	-0.66	107.50	0	6	-0.91
Citric acid	$C_6H_8O_7$	-1.72	122.70	0	8	-1.88
N,N-Diethylacetamide	$C_6H_{13}N_1O_1$	0.34	103.59	0	4	0.41
2,2-Dimethylbutane	C_6H_{14}	3.82	92.66	0	0	3.90
Acetal	$C_6H_{14}O_2$	0.60	107.44	0	4	0.53
p-Nitrobenzoic acid	$C_7H_5N_1O_4$	1.89	107.60	0	2	1.98
m-Nitrobenzamide	$C_7H_6N_2O_3$	0.77	110.44	0	4	0.62
Phenylnitromethane	$C_7H_7N_1O_2$	1.75	100.02	0	2	1.74
Cyclohepta-1,3,5-triene	C_7H_8	2.63	80.71	0	0	2.57
N-Ethylpiperidine	$C_7H_{15}N_1$	1.88	108.36	0	2	2.00
Butyltrimethylsilane	$C_7H_{18}Si_1$	4.2	128.81	0	0	4.10
Phenylethanenitrile	$C_8H_7N_1$	1.59	95.87	0	2	1.60
Phenylacetamide	$C_8H_9N_1O_1$	0.45	106.80	0	4	0.51
Ethyl nicotinate	$C_8H_9N_1O_2$	1.32	114.64	0	3	1.48
Phenetole (Ethoxybenzene)	$C_8H_{10}O_1$	2.51	102.76	0	1	2.55
1-Aminooctane	$C_8H_{19}N_1$	3.09	130.66	0	2	2.71
Allylbenzene	C_9H_{10}	3.23	105.03	0	0	3.34
3-Phenylpropionamide	$C_9H_{11}N_1O_1$	1.03	120.71	0	4	0.95
Isopropylbenzene	C_9H_{12}	3.66	109.97	0	0	3.50
Di-*tert*-butyl ketone	$C_9H_{18}O_1$	3.00	135.94	0	2	2.88
Nicotine	$C_{10}H_{14}N_2$	1.17	136.06	0	4	1.44
Phenothiazine	$C_{12}H_9N_1S_1$	3.97	144.55	0	1	3.88
Azobenzene	$C_{12}H_{10}N_2$	3.82	141.93	0	1	3.79
Phenylsulfoxide	$C_{12}H_{10}O_1S_1$	2.06	141.64	0	3	2.34
1,3-Dimethylnaphthalene	$C_{12}H_{12}$	4.42	132.42	0	0	4.21
Acetylacetone, benzal	$C_{12}H_{12}O_2$	1.59	148.28	0	4	1.83
Cyanamphenicol	$C_{12}H_{13}N_3O_5$	-0.22	194.23	0	9	-0.32
Ibuprofen	$C_{13}H_{18}O_2$	3.94	173.93	0	2	4.09
Moprolol	$C_{13}H_{21}N_1O_3$	1.69	196.84	0	6	1.93
Naproxen	$C_{14}H_{14}O_3$	3.18	174.92	0	3	3.40
Diazepam	$C_{16}H_{13}Cl_1N_2O_1$	2.85	208.44	0	5	3.02
Chlordiazepoxide (Librium)	$C_{16}H_{14}Cl_1N_3O_1$	2.44	217.04	0	6	2.57
Ampicillin	$C_{16}H_{19}N_3O_4S_1$	0.96	248.28	0	10	0.67

Table 4.1—continued

Compound	Formula	log *P*	*v*	*η*	*π*	log P_{calc}	
Triphenylene	$C_{18}H_{12}$	5.49	177.29	0	0	5.64	
Indomethacin	$C_{19}H_{16}Cl_1N_1O_4$	3.14	253.65	0	7	3.01	
Dicloxacillin	$C_{19}H_{17}Cl_2N_3O_5S_1$	2.93	310.39	0	10	2.65	
Pipamperone	$C_{21}H_{30}F_1N_3O_2$	2.40	295.99	0	10	2.19	
Tetracycline	$C_{22}H_{24}N_2O_8$	-1.47	310.03	0	16	-1.69	
Dexamethasone acetate	$C_{24}H_{31}F_1O_6$	2.91	327.33	0	10	3.19	
Fluocinonide	$C_{26}H_{32}F_2O_7$	3.19	358.78	0	12	2.75	
Large, unsubstituted compounds with connected rings; correction is suggested[a]							
Morphine	$C_{17}H_{19}N_1O_3$	0.83	213.45	0	6	1.54	*2.46*
Cocaine	$C_{17}H_{21}N_1O_4$	2.09	232.89	0	6	2.08	*3.07*
Atropine	$C_{17}H_{23}N_1O_3$	1.81	231.10	0	6	2.03	*3.02*
Pentazocine	$C_{19}H_{27}N_1O_1$	3.81	244.05	0	3	3.51	*5.60*
Testosterone	$C_{19}H_{28}O_2$	3.31	243.62	0	4	3.30	*4.86*
Progesterone	$C_{21}H_{30}O_2$	3.87	265.98	0	4	3.86	*5.57*
Tetrahydrocannabinol, Δ^1	$C_{21}H_{30}O_2$	3.78	268.83	0	2	3.65	*7.11*
Betamethasone, 17,21-dipropionate	$C_{28}H_{37}F_1O_7$	4.07	384.97	0	10	4.03	*5.02*

[a]Uncorrected predictions are in italic.

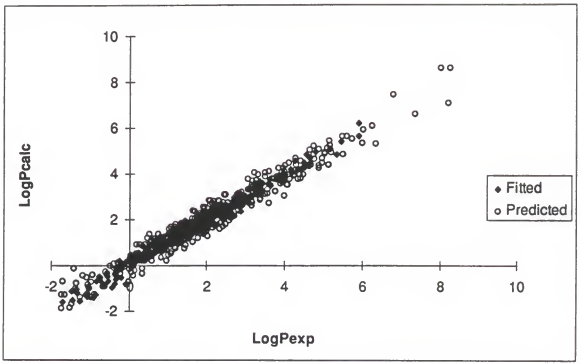


Figure 4.6 Calculated versus experimental log $P_{o/w}$ for the present model for 320 molecules where regression was performed (fitted) and 438 molecules where values were predicted with the present model.

Hydrogen Bonding?

Using an integer as predictor is certainly a disputable choice, but the similarity of residuals for different groups (figure 4.3) and the good correlations obtained for a large variety of compounds are undeniably supporting arguments for such a restrictive choice. In addition, even if we still lack a clear explanation, there is some evidence that π is related to the number of hydrogen bonds formed at the acceptor sites of the solute molecule when it is transferred from octanol to water. This is not surprising and agrees well with some of the observations derived in the many different studies attempting to connect partition coefficients to hydrogen-bonding ability and other physicochemical properties

using a methodology based on the solvatochromic model (Kamlet et al., 1988; Taft et al., 1985, 1996; El Tayar et al., 1992; Leahy et al., 1992). They suggest that besides volume, hydrogen bond acceptor basicity (β) is the most determining factor (Taft et al., 1985; Kamlet et al., 1988; El Tayar et al., 1992). This may only be true for this partitioning system, as the importance of solute hydrogen bond donor acidity (α) is diminished here by the nearly equal solvent β-strength of water and octanol (El Tayar et al., 1993; Leo, 1993; Taft et al., 1996). Although π^* has only integer values, it correlates quite well with the solvatochromic β: for the 257 compounds included in Kamlet's 1988 paper (tables II-X) we have $\beta = 0.114 + 0.118\,\pi^*$, with $r = 0.939$ and $\sigma = 0.07$. However, while the linear solvation energy relationship approach uses a β either derived from solvent effects or calculated by a rather complex set of (apparently empirical) rules with values "settled down" after a series of successive approximations, here π^* can be obtained solely from the structure using a computerized algorithm even for complex structures where there are no clear rules yet to obtain β. It also worth noting, that there seems to exist a good correlation between β and the minimum value of the electrostatic potential calculated with *ab initio* self-consistent field molecular orbital (SCF-MO) wave functions (Murray et al., 1991), and a log *P* predictive model that uses such a parameter to represent hydrogen bonding has been developed recently (Haeberlein and Brinck, 1997).

What is even more important, a change by one unit in π^* is related to an average free energy change

$$G^0_{\pi^*} = a_2 RT \ln 10 = 4.2 \text{ kJ/mol} \qquad (4.4)$$

which is in excellent agreement with the accepted Gibbs free energy for hydrogen bonds. It seems (Grant and Higuchi, 1990) that while the bonding enthalpy of hydrogen bonds varies depending on the involved systems over a relatively wide range (5-30 kJ/mol), an enthalpy-entropy compensation (Lumry and Rajender, 1970; Tomlinson, 1983), characteristic especially for water, keeps the Gibbs free energy (G) in a better defined area around 4-5 kJ/mol (1-1.2 kcal/mol), a value also obtained in recent protein stability studies (Fehrst et al., 1985; Shirley et al., 1992). Meot-Ner's observation for ionic hydrogen bonds (Meot-Ner and Sieck, 1983) might reveal why this parameter is almost unaffected by spatial crowding: "as long as there exists a single conformation in which the hydrogen bond (...) can attain optimal geometry, the bond strength is not weakened by steric crowding". These would connect π^* to real physicochemical phenomena, making it more than just an arbitrary factor.

This hypothesis is reinforced by the need to subtract one from π^*s value to get the correct result for molecules where internal hydrogen bonding is possible (e.g., carbonyl fragment ortho to OH). This indicates again a possible connection between π^* and hydrogen bond acceptor ability: the internal occupancy of one hydrogen bond acceptor site in the solute causes exactly the expected decrease of 1 in π^*. The corresponding correction in the predicted log P ($1a_2 = 0.73$) is

quite close to that measured between, for example, *o*- versus *m*-hydroxybenzoic acid (2.24 - 1.49 = 0.75) or even to that used by fragmental methods for intramolecular H bonding: CLOGP (0.63) (Leo, 1993), Rekker's revised method (0.66) (Mannhold et al., 1995). This correction to α is automatically performed when OH and carbonyl groups are detected in the ortho position. The fact that the same α parameter can account for corrections due to some electron-withdrawing groups and due to intramolecular hydrogen bonding represents not a weakness but a strength of the model. It proves again that this parameter is related to hydrogen bond acceptor basicity, which is clearly diminished by such substitutions or intramolecular hydrogen bonds.

This connection could explain why coupling a continuous variable (volume, for cavity formation) with a parameter having only integer values (modeling H-bond formation) can lead to an oversimplified but reasonably good model, which also agrees well with the widely accepted partially ordered structure of water. The model is unquestionably far too simple to account for all effects influencing partition and solubility, but it might represent a framework for future study.

Comparison with Other Methods

The present method assumes the contribution of hydrogen atoms to log *P* to be proportional to their contribution to molecular volume. Considering an average of this volume as sliced off by a plane from the rest of the molecule, this is quite small, being around $a_1 v_{Hfrg} \approx 0.09$. This perspective reveals a major

problem of CLOGP's "constructionist approach": the contribution of a hydrogen atom to molecular volume or surface area in usual organic molecules (where it is bonded to heavier, larger atoms) is only part of what it is in a hydrogen molecule, around 60-70% for volume and 70-75% for surface. Thus, however accurately log *P* is measured for H_2, by using half of it as hydrogen contribution CLOGP uses an overestimate for its very "fundamental" chemical fragment. This might very well be the main reason why a slope appreciably less than 1.0 is obtained (Leo, 1993) when regressing the measured partition coefficients against CLOGP, even if strong deviants are removed. Indeed, in Rekker's method – which uses the same basic fragmental approach as CLOGP – the value obtained (Rekker and de Kort, 1979) for f_H with multiple regression in a statistical approach is about 20% smaller: 0.182 versus 0.227. A 2D model will almost inevitably overweight the role of hydrogen fragments due to their large frequency. Assigning a too large value for them will require not only more negative contributions for the hydrophilic groups but also some otherwise probably unneeded correction factors to compensate for the overestimate. Even in the Rekker methodology individual H fragments are included whenever possible in larger fragments to avoid the need for branching and proximity corrections. In a recently published (Meylan and Howard, 1995) variation of the same "two dimensional" approach (AFC model), hydrogen is not even considered explicitly, only as part of other fragments. ALOGP's trend (Viswanadhan et al., 1989) differs considerably from all others' probably due to

the unrealistic distinction used for hydrogen atoms in this model (10 different H fragment constants).

Another reason for the differences is the way in which alkanes are treated. The AFC model, developed on a large database, cannot give a good fit for them, while both Rekker's method and CLOGP were mainly built starting from these smaller molecules. The peculiar behavior of the alkanes may be another likely reason to overweight H fragments.

Rekker's "magic constant" made room for much controversy, but the fact that a whole model could be built around it is some sort of evidence for a "quantified" behavior in log *P* even if the value of the constant is too small and the rules for choosing its multiplicator, the "key number," are not always sufficiently clear to represent a definite proof. As the present method also uses a statistical approach and a parameter with integer values, it is probably of interest to mention that here most polar groups, disregarding their volume contribution, introduce a change of $2a_2 = 1.45$, which is exactly 5 times Rekker's $C_M = 0.29$. Considering that both values are statistically very well defined, this agreement is striking, but it might as well be purely coincidental, as C_M has a quite small value that even has been redefined lately (Rekker and Mannhold, 1992; Mannhold et al., 1995). Nevertheless, it might be an explanation for the "magic" reoccurrence of such values. While here for most fragments the difference between alkyl and aromatic contribution is $1a_2 = 0.726$ (being mainly related to the change by one

unit in π, the fragment's volume is usually only slightly affected), Rekker's difference is usually 3 times his "magical" constant: $3C_M = 0.87$.

Starting from more realistic 3D structures, the present model does not use hundreds of empirically derived fragment contributions and is much simpler than all these methods, yet in most cases it performs no worse than these complicated methods. In a comparison for 48 drugs (antiarrhythmics, β-blockers, phenothiazines, and benzamides) (Table 4.2) used in a study by Mannhold et al. (1990), the correlations show a performance close to that of CLOGP and Rekker's method. If mean squared errors (MSE) relative to the experimental data are considered, the present model actually performs slightly better than both since even if Σf values (Rekker) correlate best, they systematically overestimate log *P* on this set:

All 48	**Present**	**CLOGP**	**Rekker**
r	0.929	0.926	0.947
σ	0.57	0.70	0.50
MSE	0.57	0.71	0.72

Even for the set of test compounds used in a very recent comparison (Mannhold et al., 1995) of predictive power for molecular lipophilicity, where two groups of methods were identified with differing predictive power (CLOGP and revised Rekker versus ALOGP and Suzuki/Kudo), the present method's performance would place it in the first, better performing group. We present here only the statistical data since the 65 simple organic molecules used were mainly

part of our original data set, and the 40 more complicated drug molecules included in their table are basically a subset of the previous 48 (Figure 4.7).

65 Simple	**Present**	**CLOGP**	**Rekker**	**ALOGP**	**Suzuki/Kudo**
r	0.986	0.988	0.987	0.976	0.974
σ	0.19	0.19	0.21	0.24	0.27
MSE	0.21	0.20	0.23	0.30	0.30
40 Drugs	**Present**	**CLOGP**	**Rekker**	**ALOGP**	**Suzuki/Kudo**
r	0.940	0.924	0.960	0.896	0.898
σ	0.52	0.72	0.46	0.52	0.62
MSE	0.54	0.71	0.45	0.90	1.12

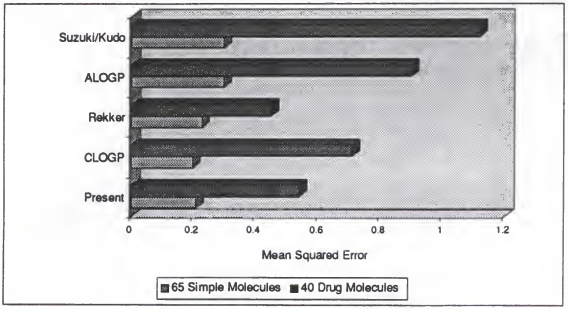


Figure 4.7 Comparison of molecular lipophilicity predictive power for different models on a set of 65 simple and 40 drug molecules (antiarrhytmics, β-blockers, phenotiazines, benzamides) from a recently published article (Mannhold et al., 1995). Smaller mean squared errors (MSE) represent better agreements with experimental values.

Table 4.2. Experimental and Predicted Partition Coefficients for 48 Drugs Used by Mannhold et al. (1990).

Compound	log P_{exp}	Present	CLogP	Rekker
Aprindine	4.86	6.74	4.77	5.86
Asocainol	4.85	5.30[a]	6.23	6.66
Carocainide	1.40	2.11	2.12	2.57
Disopyramide	2.58	3.22	1.35	2.51
Flecainide	3.78	3.46	4.22	4.41
Indecainide	3.11	2.57[a]	2.46	3.33
Lidocaine	2.42	2.79	1.66	3.04
Lorcainide	4.85	5.14	4.64	5.25
Mexiletine	2.15	2.76	2.87	3.15
Nicainoprol	1.63	1.30	1.53	2.05
Procainamide	0.82	1.92	1.11	1.21
Quinacainol	3.67	4.49	3.68	4.37
Quinidine	3.71	2.16[a]	2.74	3.36
Ethmozine	2.98	2.01	2.72	3.06
Propafenone	4.63	3.82	3.36	4.31
Acebutolol	1.71	2.15	1.61	2.40
Alprenolol	3.13	3.15	2.74	3.58
Atenolol	0.78	0.33	-0.11	0.75
Bunitrolol	1.60	2.15	1.74	2.09
Bupranolol	2.71	3.40	3.20	3.71
Metipranolol	2.66	3.02	3.13	4.06
Metoprolol	1.98	2.12	1.35	2.35
Oxprenolol	2.29	2.65	1.62	3.14
Penbutolol	3.92	4.36	4.19	4.68
Pindolol	1.97	2.01	1.65	2.43
Propranolol	3.17	3.12	2.75	3.66
Alimemazine	4.81	4.73	4.59	5.05
Chlorpromazine	5.34	4.83	5.20	5.27
Fluphenazine	4.79	4.22	5.90	5.56
Levomepromazine	4.86	4.67	4.64	5.13
Perazine	4.00	4.23	5.22	4.53
Perphenazine	4.47	4.00	5.57	5.16
Promazine	4.64	4.29	4.28	4.53
Promethazine	4.75	4.29	4.65	5.11
Sulforidazine	4.45	4.39	4.71	4.37
Thiethylperazine	5.18	5.65	6.31	6.12
Thioridazine	5.99	6.33	6.42	6.35
Trifluperazine	5.07	5.00	6.48	6.12
Triflupromazine	5.35	5.06	5.53	5.67
Alizapride	1.79	1.82	2.19	2.91
Alpiropride	1.69	0.82	0.42	1.05
Amisulpride	1.10	1.49	1.05	1.48
Bromopride	2.96	2.60	2.04	3.09
Metoclopramide	2.91	2.39	1.89	2.90
Sulpiride	0.42	0.53	0.95	1.68
Sultopride	1.06	1.94	1.74	2.50
Tiapride	1.02	1.31	0.76	1.83
Veralipride	1.47	0.75	0.41	1.92

[a]Obtained using the correction for large, rigid, unsubstituted molecules. The program's present version suggested this correction in these three cases, though is certainly not needed for quinidine.

Testing of the program on the set of 35 nucleosides and 12 nucleoside bases of Viswanadhan et al. (1993) revealed the necessity to use a correction of one in π for every hydroxy group present in the riboside ring indicating possible, but not thoroughly justified intramolecular hydrogen bonds in these nucleosides. With this correction, results are even better than those of the models used in the original study as shown here by the correlation coefficients and mean squared errors for the bases (no correction used) and the whole set:

Present	ALOGP	CLOGP	BLOGP
Bases (12):			
0.791 (0.47)	0.797 (0.40)	0.596 (0.60)	0.810 (0.51)
All (47):			
0.873 (0.40)	0.842 (0.51)	0.711 (0.92)	0.398 (1.19)

By avoiding overparametrization, the model seems to have a promising behavior for large molecules, like nonzwitterionic peptides. This is subject to further study since the intramolecular hydrogen bonding may cause some problems, but let us give here just one example to prove the advantage of such an approach over traditional 2D fragment ones: The present model predicts a partition coefficient of 1.28 for cyclosporin-A, an immunosuppressive cyclic undecapeptide of fungal origin far larger than any molecule included in the training set. Its log *P* has been estimated (Cefalu and Pardridge, 1985; El Tayar et al., 1993) to be 2.92-2.99. However, HPLC determination (Takada et al., 1992) suggests a log *P* = 0.55. The difference may be related to the possibility of three internal H-bonds (El Tayar et al., 1993) – causing an increase of 2.2 in log *P* according to the present method, quite close to the actual difference of 2.4 –

or just to the experimental difficulty of determining log *P* for such a large compound, but anyway the present prediction is orders of magnitude better than CLOGP's unrealistic 14.0 (El Tayar et al., 1993).

The model developed correlates successfully molecular size and log *P* via a quantified parameter determined by the functional groups present in the solute molecule, a parameter describing a free energy change that agrees well with that accepted for hydrogen bonds in water. The remarkable economy in parameters clearly distinguishes the present model from usual fragment-based models, yet the model works well enough even on pharmaceutically important molecules. It gives some insight into the behavior of "two dimensional" fragment approaches and represents a possible framework for future study. A corresponding computer program (Figure 4.8) has been developed.

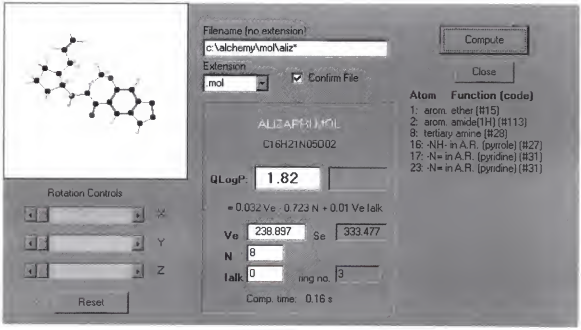


Figure 4.8 Interface of the current version of the log *P* computing program.

CHAPTER 5
NONZWITTERIONIC PEPTIDES: OCTANOL-WATER PARTITION

Background

As discussed in the previous chapter the partition coefficient between *n*-octanol and water ($P_{o/w}$) is defined for dilute solutions as the molar concentration ratio of a single species (A) between two phases at equilibrium,

$$P = \frac{[A]_o}{[A]_w} \qquad (5.1)$$

and its logarithm (log $P_{o/w}$) is an important parameter widely used to characterize molecular lipophilicity and hydrophobicity. Despite this, an area of increasing pharmaceutical interest – that of amino acids and peptides – is still lacking general, fast, and relatively reliable predictive methods. Most models developed for other molecules become quite unreliable here, while methods specifically developed for peptides cannot be used if nonpeptide constituents are present. A predictive method might prove useful here not only because many, already established relationships relying on log *P* might prove useful in peptide drug delivery, dosage formulation, and so on, but also because the hydrophobicity of the component amino acids is a key to the final three-dimensional structure and biological activity of peptides, proteins, and enzymes. Hydrophobicity as measured by octanol-water partition seems to be again one of the best

predictors. For the biological transport properties of peptides log P_{ow} sometimes is a good predictor (Banks and Kastin, 1985; Hunter et al., 1990; Samii et al., 1994), but sometimes it is not as good as it is for other solutes (Burton et al., 1991, 1992; Begley, 1992; Conradi et al., 1992; Chikhale et al., 1994). Considering the peculiar nature of peptides (conformational flexibility, presence of ionizable species, possible multiple intramolecular hydrogen bonds, active transport mechanisms) this is understandable; some aspects have been reviewed (Burton et al., 1991), but many reasons still await elucidation. In addition, ionization-related problems are rarely rigorously considered.

We will adopt the usual convention, that is, reserve the term *partition coefficient* (*P*) for distribution of neutral species, and designate the values measured at a given pH as *distribution coefficient* (*D*), or *apparent partition coefficients*. Ion-pair partitioning and the extraction constant formalism will be briefly reviewed to describe the sigmoidal lipophilicity profile of ionizable, nonzwitterionic peptides. It seems possible to extend the present model to estimate apparent partition coefficients measured around neutral pH and physiological conditions for monoionic peptides; however, as no standard conditions are yet defined and only relatively small number of experimental data are available, the situation here is more complex.

Compared to the abundance of partition data for other compounds, only a relatively small number of experimental data are available for peptides. Most of these compounds are hydrophilic, and the reliable determination of the resulting

small concentrations in the nonaqueous phase requires sufficiently sensitive methods. When ionizable residues are present, the values become pH- and buffer-dependent, and consequently quite inconsistent among studies. However, the pharmaceutical interest of the last few years resulted in an increasing amount of data (Yunger and Cramer, 1981; Fauchère and Pliska, 1983; Itokawa et al., 1984; Cefalu and Pardridge, 1985; Cotton et al., 1986; Bock et al., 1988; Bundgaard and Moss, 1990; Hunter et al., 1990; Tokumura et al., 1990; Rosenberg et al., 1990; Burton et al., 1991, 1992; Akamatsu and Fujita, 1992; Conradi et al., 1992; Kim and Szoka, 1992; Takada et al., 1992; Kahns et al., 1993; El Tayar et al., 1993; Akamatsu et al., 1994; Chikhale et al., 1994; Katayama et al., 1994; Lan et al., 1994; Hansch et al., 1995; Wimley et al., 1996).

Because of our involvement in brain-targeted peptide delivery (Bodor et al., 1992, Prokai et al., 1994; Prokai-Tatrai et al., 1996), using a technique designated as molecular packaging (Bodor and Prokai, 1995), we became especially interested in developing a sufficiently fast, general, physicochemical based method to predict partition properties of peptides and their different chemical derivatives. The model we devised by using a molecular size-based approach (Bodor and Buchwald, 1997a) to estimate log $P_{o/w}$ for organic molecules has a promising performance in predicting the partition of 188 neutral, nonzwitterionic peptides with available experimental data. With some extensions, it even gives reasonable estimates for the apparent partition

coefficient of monoionic peptides measured at neutral pH. Part of the result confirmed again that the parameter (γ) introduced in the model is connected to hydrogen-bonding ability.

Experimental Data

Experimental data used are from the literature cited above; the average value was used if more acceptable measurements were available. Regressions were performed with a standard spreadsheet program (Microsoft Excel 5.0). For sandostatin (octreotide) octanol-water partition coefficients were determined by using shake flask method according to standard requirements and recommendations. Commercially available pH 7.00 potassium phosphate monobasic sodium hydroxide buffer (FisherChemical, 0.05 M) was used diluted to 0.01 M. The ionic strength was adjusted to 0.1 with NaCl. For the neutral form 0.1 M NaOH buffer was used. Both solvents (octanol and aqueous) were presaturated with the other by leaving them in contact for at least 24 hours. Initial solute concentrations were as low as possible around 0.1 mM. To 3 ml of solution a known ratio of the other solvent was added in a Teflon-capped glass tube, and the mixture was vigorously shaken for long enough to reach equilibrium (more than 200 inversions, more than 10 minutes). After equilibration for a few minutes, the mixture was centrifuged (10 minutes at 2,000 rpm) to achieve good separation. Final concentrations in both phases were determined by absorption measurements, using both UV spectrometry and HPLC procedures. The partition coefficient was computed from the concentration ratios

estimated from the absorption measurements correcting for unequal solvent volumes when they were used, for example, if aqueous concentrations were measured

$$P = (c_{in} - c_{fin}) / c_{fin} \times V_w / V_o \qquad (5.2)$$

The volume ratios were selected so to be close to equidistribution of compound in the two phases, and for each determination at least two different ratios were used. The UV spectrum between 200 and 400 nm was recorded with a Perkin Elmer Lambda 11 UV/VIS spectrometer connected to a computer running the PECCS software package. Concentrations have been evaluated at different wavelengths around the absorption maxima above 250 nm, and the obtained average was used. In addition, concentration measurements based on comparison of peak areas from HPLC procedures were also used. The HPLC system consisted from a Spectra Physics (San Jose, CA, USA) SP8810 precision isocratic pump, SP8450 variable-wavelength UV-Vis detector, SP4290 computing integrator, and SP4290 autosampler. The column used was a 15-cm Altima C_{18} analytical column with 5-µm particle-size packing; the mobile phase was 40% acetonitrile, 60% buffer (pH 4, 5 mM ammonium acetate with acetic acid). Flow rate was 0.5-1 ml/min with an injection loop of 20 µl, and detection at 280 nm.

Partition coefficients were estimated with the fully computerized model already described. The corresponding program requires as input a reasonable 3D molecular structure; here they were obtained with a simple molecular

mechanics optimization (minimize procedure from Alchemy II program; Tripos Assoc., St. Louis, MO).

Model

The main equation used is

$$\log P = 0.032\, v - 0.723\, \pi \qquad (5.3)$$

It was obtained, as described in the previous chapter, after regression on a set of 320 different molecules ($r = 0.989$, $\sigma = 0.214$), and after implementation of the corresponding computer program its predictive power was tested on an additional set of 438 molecules ranging from hydrogen to prednisolone, sulfanilamide, or propicillin ($r = 0.975$, $\sigma = 0.365$) (Bodor and Buchwald, 1997a).

The π contributions are assigned by a fully automated algorithm, and for the usual heteroatom functions present in peptides they are as follows: 4 for -CON<; 3 for peptide -CONH-, $-CONH_2$; 2 for $-NH_2$, -NH-, -N<, -COOH, -COO- - OH; 1 for Ar-OH, -S-, Ar(-N=), Ar(-NH-); 0 for -SH, -S-S-. Here Ar represents an aromatic substitution, and Ar() an atom included in an aromatic ring structure. The contribution of each function is fixed, and for most molecules the assumption of a simple additivity works reasonably well. The rules are sufficiently clear and consistent to be easily implemented in a computer program. Using an integer as predictor is a disputable but very restrictive choice. It is worth noting that, even if a clear explanation is still lacking, there is some

evidence that $\mathcal{H}$ is related to the hydrogen bonds formed at the acceptor sites of the solute molecule when it is transferred from octanol to water.

Two reasonable modifications had to be made. For structures containing side chains with aliphatic -OH groups (Ser, Thr) log *P* was consistently underestimated with about 0.5-0.7 units. This occurred for practically all cases, and the value of the difference indicated that it might be the result of an intramolecular hydrogen bond resembling the situation of ortho -OH and -CO- substituents on aromatic rings. Therefore, in a similar manner, a correction of $\Delta\mathcal{H}$ = +1, thus Δlog P_{pred} = +0.72, was automatically included when an intramolecular six-membered ring can be formed through H-bonding between a hydroxyl and carbonyl function. A very similar error occurred for hydrogen-bearing amide functions present in the backbone or in the side chains (Gln, Asn), therefore, the same $\Delta\mathcal{H}$= +1 correction was included for peptide -CONH- and -$CONH_2$ groups diminishing their $\mathcal{H}$ contribution from 4 to 3 as indicated in the above paragraph. Hydrogen bonding in these structures was hypothesized by others as well, for example, to explain the high coefficient values for Ser, Thr, Asn, and Gln in the model of Akamatsu et al. (1994) or to explain the discrepancies between measured partition coefficients and CLOGP calculated values (Abraham and Leo, 1987). In addition, there is a possibility that such intramolecular bonding occurs not only directly but also through water-bridging (Abraham and Leo, 1987). The fact that the present model can account for these shifts via a change by one unit in the same $\mathcal{H}$ parameter can be considered an additional strength of

the model as it reveals again the connection between this parameter and hydrogen bond acceptor ability.

The amide function is by all accounts a problematic one regarding its contribution to the partition coefficient. Even the fragment values used in different fragmental methods cover a wide range: from -0.524 (AFC – Meylan and Howard, 1995) to -2.180 (CLOGP – Hansch and Leo, 1979). This is much wider than for other fragments, for example, for -COOH this range is from -0.690 (AFC) to -1.110 (CLOGP), while for $-NH_2$ from -1.415 to -1.540. As it contains two different heteroatoms and is both hydrogen-bond donor and acceptor, it is unlikely that a single fragment value can describe its behavior. Indeed, most methods have to use many, quite large, additional correction factors.

Partition Coefficients

By definition, partition coefficients describe the partitioning of neutral compounds. The compounds included here are mainly peptides with blocked N and C termini (e.g., N-acetyl peptide amides). For those having un-ionizable side chains the measured partition coefficients are only very slightly pH-dependent, and most data can be used without any special problem; partition coefficients measured at usual pH values are those of the neutral species. The set also includes data for compounds with acidic or basic ionizable side chains if they were measured at pH values where their neutral form is prevalent (Asp, Glu: $pH \leq 2$; Arg, Lys, Orn: $pH \geq 13$; His: $pH \geq 10$). As histidine residues contain an imidazole group with a pK_a in the 6.5-7.0 range (Perrin, 1965), apparent partition

coefficients of the corresponding compounds are pH-dependent. Values measured at neutral pH are only slightly smaller (Kim and Szoka, 1992; Akamatsu et al., 1994) than those of the neutral forms; therefore, if only values measured at pH = 7 or 7.4 were available they were also included here.

For the 188 peptides where partition data for the neutral form were available (table 5.1, figure 5.1):

$$\log P = -0.166(\pm 0.036) + 1.066(\pm 0.023) \log P_{pred}$$
$$n = 188;\ r^2 = 0.914;\ \sigma = 0.455;\ F = 1978.0 \qquad (5.4)$$

This is a good correlation for such a remarkably simple model that uses only two parameters and was not specifically developed for peptides – especially considering that many compounds here are significantly larger than those used to develop the model.

As a comparison, a recently released variation of the same 2D fragment method (ACD/LogP), which uses basic fragment values very similar to those of CLOGP, despite using over 500 different "fragmental contributions" and over 2000 different "intermolecular interaction contributions," not to mention the "atomic contributions" of the different carbon atoms, gives for the same data

$$\log P = +0.452(\pm 0.035) + 0.867(\pm 0.021) \log P_{ACD}$$
$$n = 188;\ r^2 = 0.904;\ \sigma = 0.481;\ F = 1750.2 \qquad (5.5)$$

The correlation itself is not much worse, but contrary to the previous equation here both the intercept and the slope are significantly different from what they should be (zero, respectively one), therefore, the actual error of

prediction as revealed by the mean squared errors (0.698 vs 0.480) is clearly higher.

The difference also becomes more obvious if a zero intercept, required for a correct scaling, is imposed on the correlations. This produces only a very small change for the present model, but a significant worsening in the correlation for ACD/LogP:

$$\log P = 0 + 0.968(\pm 0.022) \log P_{pred}$$
$$n = 188;\ r^2 = 0.904;\ \sigma = 0.480;\ F = 1762.4 \qquad (5.6)$$

respectively

$$\log P = 0 + 0.862(\pm 0.028) \log P_{ACD}$$
$$n = 188;\ r^2 = 0.818;\ \sigma = 0.659;\ F = 841.3 \qquad (5.7)$$

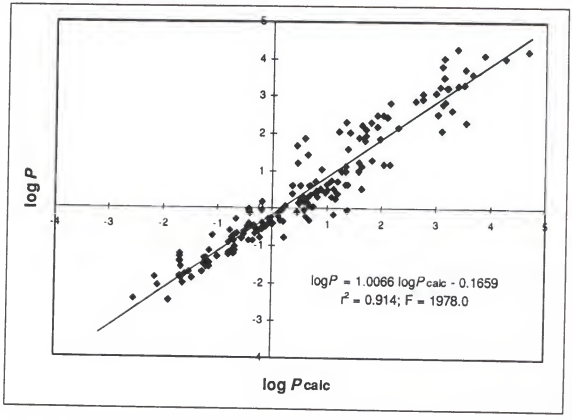


Figure 5.1 Experimental versus predicted log *P* for 188 neutral peptides.

The present method not only achieves this performance with a remarkably small number of parameters, but, for three large, cyclic peptides with available experimental log *P* determined by shake-flask methods, cyclosporin (Cefalu and Pardridge, 1985; Takada et al., 1992; El Tayar et al., 1993), melanotan-II (Lan et al., 1994), and sandostatin, it also performs several orders of magnitude better than multiparameter 2D fragment methods (table 5.1). These peptides were not included in the above correlations and will be discussed later. While the present method gives reasonable predictions (within one or two log units) for these large compounds, CLOGP and ACD/LogP – probably due to overparametrization – give a few unrealistic ones (table 5.1). This represents one of the most promising feature of the present model: by avoiding overparametrization its predictions for large molecules seem more realistic than those obtained from multiparameter fragment methods.

It is also worth noting that the present model gives better predictions than 2D fragment methods for the set of 16 basic amino acids with blocked termini (Ac-X-NH_2), not considering the ionizable aspartate, glutamate, arginine, and lysine where data was not available for the neutral forms: $r^2 = 0.902$, $\sigma = 0.283$, $F = 129.5$, while for CLOGP (Abraham and Leo, 1985) $r^2 = 0.843$, $\sigma = 0.359$, $F = 75.2$, respectively, for ACD/LogP $r^2 = 0.823$, $\sigma = 0.382$, $F = 64.9$. With the present method the biggest errors occur for phenylalanine and tyrosine; both are overestimated, and this tendency seems to be present for most larger peptides containing these residues.

Table 5.1 Octanol-Water Partition Data for the Peptides Included in the Present Study.

Name	Formula	log P	log D^a	v (Å^3)	π	log P_{pr}	log D_{pr}[b]	log P_{ACD}
AcPheNH$_2$ (F)	$C_{11}H_{14}N_2O_2$	0.04	0.04	159.14	6	0.73	0.73	-0.13
AcIleNH$_2$ (I)	$C_8H_{16}N_2O_2$	-0.03	-0.03	141.96	6	0.18	0.18	-0.60
AcLeuNH$_2$ (L)	$C_8H_{16}N_2O_2$	-0.04	-0.04	142.07	6	0.19	0.19	-0.60
AcValNH$_2$ (V)	$C_7H_{14}N_2O_2$	-0.61	-0.61	127.67	6	-0.27	-0.27	-1.13
AcTyrNH$_2$ (Y)	$C_{11}H_{14}N_2O_3$	-0.79	-0.79	165.96	7	0.22	0.22	-0.87
AcAlaNH$_2$ (A)	$C_5H_{10}N_2O_2$	-1.47	-1.47	100.15	6	-1.15	-1.15	-2.01
AcGlyNH$_2$ (G)	$C_4H_8N_2O_2$	-1.76	-1.76	86.01	6	-1.60	-1.60	-2.36
AcProNH$_2$ (P)	$C_7H_{12}N_2O_2$	-1.34	-1.34	120.06	7	-1.24	-1.24	-0.48
AcTrpNH$_2$ (W)	$C_{13}H_{15}N_3O_2$	0.42	0.42	183.26	7	0.77	0.77	-0.21
AcSerNH$_2$ (S)	$C_5H_{10}N_2O_3$	-1.87	-1.87	106.80	7	-1.66	-1.66	-2.33
AcThrNH$_2$ (T)	$C_6H_{12}N_2O_3$	-1.57	-1.57	120.91	7	-1.21	-1.21	-1.98
AcMetNH$_2$ (M)	$C_7H_{14}N_2O_2S_1$	-0.47	-0.47	144.68	7	-0.45	-0.45	-0.96
AcHisNH$_2$ (H)	$C_8H_{12}N_4O_2$	-1.70	-1.70	141.52	8	-1.28	-1.28	-2.43
AcCysNH$_2$ (C)	$C_5H_{10}N_2O_2S_1$	-0.29	-0.29	115.33	6	-0.67	-0.67	-1.12
AcGlnNH$_2$ (Q)	$C_7H_{13}N_3O_3$	-2.05	-2.05	138.59	9	-2.09	-2.09	-2.94
AcAsnNH$_2$ (N)	$C_6H_{11}N_3O_3$	-2.41	-2.41	124.64	9	-2.54	-2.54	-2.84
AcGluNH$_2$ (E)	$C_7H_{12}N_2O_4$		-2.47	135.77	8	-1.46	-3.63	-1.99
AcAspNH$_2$ (D)	$C_6H_{10}N_2O_4$		-2.60	121.89	8	-1.90	-4.07	-1.89
AcArgNH$_2$ (R)	$C_8H_{17}N_5O_2$		-2.84	165.29	11	-2.69	-3.41	-3.11
AcLysNH$_2$ (K)	$C_8H_{17}N_3O_2$		-2.82	152.32	8	-0.93	-3.10	-2.37
AcAdamAlaNH$_2$	$C_{15}H_{24}N_2O_2$	1.90	1.90	214.89	6	1.58	1.58	1.40
AcCystineNH$_2$	$C_{10}H_{18}N_4O_4S_2$	-1.70	-1.70	224.70	12	-1.52	-1.52	-1.44
AcGly(dimethyl)-NH$_2$	$C_6H_{12}N_2O_2$	-0.87	-0.87	114.02	6	-0.71	-0.71	-1.66
AcGly(diallyl)-NH$_2$	$C_{10}H_{16}N_2O_2$	0.05	0.05	159.55	6	0.74	0.74	0.09
AcGly(dipropyl)-NH$_2$	$C_{10}H_{20}N_2O_2$	0.75	0.75	170.09	6	1.08	1.08	0.46
AcGly(dimethallyl)-NH$_2$	$C_{12}H_{20}N_2O_2$	1.12	1.12	187.47	6	1.63	1.63	1.19
AcGly(diisobutyl)-NH$_2$	$C_{12}H_{24}N_2O_2$	1.87	1.87	197.95	6	1.96	1.96	1.16
AcGly(dibenzyl)-NH$_2$	$C_{18}H_{20}N_2O_2$	2.09	2.09	232.99	6	3.08	3.08	2.05
AcGly(allyl)-NH$_2$	$C_7H_{12}N_2O_2$	-0.86	-0.86	123.01	6	-0.42	-0.42	-1.21
Ac-Phe-NH-Me	$C_{12}H_{16}N_2O_2$	0.40	0.40	173.71	6	1.19	1.19	0.10
Ac-Leu-NH-Me	$C_9H_{18}N_2O_2$	0.14	0.14	156.45	6	0.64	0.64	-0.42
Ac-Ile-NH-Me	$C_9H_{18}N_2O_2$	0.14	0.14	156.45	6	0.64	0.64	-0.42
Ac-Tyr-NH-Me	$C_{12}H_{16}N_2O_3$	-0.32	-0.32	180.24	7	0.68	0.68	-0.64
Ac-Val-NH-Me	$C_8H_{16}N_2O_2$	-0.35	-0.35	142.40	6	0.20	0.20	-0.95
Ac-Pro-NH-Me	$C_8H_{14}N_2O_2$	-1.19	-1.19	134.05	7	-0.79	-0.79	-0.29
Ac-Ala-NH-Me	$C_6H_{12}N_2O_2$	-1.21	-1.21	114.27	6	-0.70	-0.70	-1.83
Ac-Gly-NH-Me	$C_5H_{10}N_2O_2$	-1.56	-1.56	100.30	6	-1.14	-1.14	-2.18
AcGlyValNH$_2$	$C_9H_{17}N_3O_3$	-1.33	-1.33	166.80	9	-1.19	-1.19	-1.96
AcAlaValNH$_2$	$C_{10}H_{18}N_3O_3$	-1.13	-1.13	181.07	9	-0.74	-0.74	-1.61
AcLeuValNH$_2$	$C_{13}H_{25}N_3O_3$	0.26	0.26	223.08	9	0.60	0.60	-0.20
AcGlyPheNH$_2$	$C_{13}H_{17}N_3O_3$	-0.56	-0.56	198.56	9	-0.18	-0.18	-0.96
AcIleValNH$_2$	$C_{13}H_{25}N_3O_3$	0.16	0.16	222.91	9	0.59	0.59	-0.20
AcValValNH$_2$	$C_{12}H_{23}N_3O_3$	-0.32	-0.32	208.92	9	0.15	0.15	-0.73
AcPheValNH$_2$	$C_{18}H_{23}N_3O_3$	0.43	0.43	240.97	9	1.16	1.16	0.32
AcAlaLeuNH$_2$	$C_{11}H_{21}N_3O_3$	-0.54	-0.54	194.82	9	-0.30	-0.30	-1.08
AcAlaAlaNH$_2$	$C_8H_{15}N_3O_3$	-2.00	-2.00	152.99	9	-1.63	-1.63	-2.49
AcGlyLeuNH$_2$	$C_{10}H_{19}N_3O_3$	-0.78	-0.78	180.94	9	-0.75	-0.75	-1.43

Table 5.1—continued

Name	Formula	log *P*	log D^a	*v* (Å^2)	π	log P_{pr}	log D_{pr}[b]	log P_{ACD}
AcLeuIleNH$_2$	$C_{14}H_{27}N_3O_3$	0.68	0.68	236.89	9	1.03	1.03	0.33
AcPheGlyNH$_2$	$C_{13}H_{17}N_3O_3$	-0.50	-0.50	198.38	9	-0.19	-0.19	-0.91
AcValAlaNH$_2$	$C_{10}H_{19}N_3O_3$	-1.14	-1.14	181.07	9	-0.74	-0.74	-1.61
AcTyrValNH$_2$	$C_{16}H_{23}N_3O_4$	-0.20	-0.20	246.13	10	0.61	0.61	-0.42
AcTyrLeuNH$_2$	$C_{17}H_{25}N_3O_4$	0.32	0.32	261.01	10	1.08	1.08	0.11
AcTyrPheNH$_2$	$C_{20}H_{23}N_3O_4$	0.54	0.54	278.59	10	1.64	1.64	0.59
AcTrpValNH$_2$	$C_{18}H_{24}N_4O_3$	0.73	0.73	264.30	10	1.18	1.18	0.24
AcMetValNH$_2$	$C_{12}H_{23}N_3O_3S_1$	-0.28	-0.28	225.79	10	-0.04	-0.04	-0.56
AcMetPheNH$_2$	$C_{16}H_{23}N_3O_3S_1$	0.42	0.42	257.37	10	0.96	0.96	0.44
AcSerValNH$_2$	$C_{10}H_{19}N_3O_4$	-1.53	-1.53	188.10	10	-1.24	-1.24	-1.92
AcSerPheNH$_2$	$C_{14}H_{19}N_3O_4$	-0.79	-0.79	219.67	10	-0.24	-0.24	-0.92
AcThrValNH$_2$	$C_{11}H_{21}N_3O_4$	-1.25	-1.25	202.19	10	-0.79	-0.79	-1.58
AcThrIleNH$_2$	$C_{12}H_{23}N_3O_4$	-0.86	-0.86	215.82	10	-0.36	-0.36	-1.04
AcAsnValNH$_2$	$C_{11}H_{20}N_4O_4$	-1.85	-1.85	205.39	12	-2.13	-2.13	-2.44
AcAsnIleNH$_2$	$C_{12}H_{22}N_4O_4$	-1.43	-1.43	219.38	12	-1.69	-1.69	-1.91
AcAsnPheNH$_2$	$C_{15}H_{20}N_4O_4$	-1.14	-1.14	236.49	12	-1.14	-1.14	-1.44
AcLeuAsnNH$_2$	$C_{12}H_{22}N_4O_4$	-1.30	-1.30	219.65	12	-1.68	-1.68	-1.91
AcIleAsnNH$_2$	$C_{12}H_{22}N_4O_4$	-1.41	-1.41	219.69	12	-1.68	-1.68	-1.91
AcGlnValNH$_2$	$C_{12}H_{22}N_4O_4$	-1.85	-1.85	219.37	12	-1.69	-1.69	-2.54
AcGlnLeuNH$_2$	$C_{13}H_{24}N_4O_4$	-1.32	-1.32	233.49	12	-1.24	-1.24	-2.01
AcGlnPheNH$_2$	$C_{16}H_{22}N_4O_4$	-1.14	-1.14	250.84	12	-0.69	-0.69	-1.53
AcPheGlnNH$_2$	$C_{16}H_{22}N_4O_4$	-1.03	-1.03	250.92	12	-0.69	-0.69	-1.49
AcValGlnNH$_2$	$C_{12}H_{22}N_4O_4$	-1.82	-1.82	219.45	12	-1.69	-1.69	-2.54
AcLysLeuNH$_2$	$C_{14}H_{26}N_4O_3$	-0.26		247.42	11	-0.08	-2.24	-1.44
AcLysPheNH$_2$	$C_{17}H_{24}N_4O_3$	0.12	-2.43	265.00	11	0.48	-1.68	-0.97
AcPheLysNH$_2$	$C_{17}H_{24}N_4O_3$	0.14	-2.23	264.46	11	0.47	-1.70	-0.92
AcOrnPheNH$_2$	$C_{16}H_{24}N_4O_3$	-0.41	-2.23	250.44	11	0.02	-2.15	-0.89
AcOrnLeuNH$_2$	$C_{13}H_{26}N_4O_3$	-0.71		233.43	11	-0.52	-2.69	-1.36
AcPhe-PheNH$_2$	$C_{20}H_{23}N_3O_3$	1.19	1.19	271.91	9	2.15	2.15	1.32
Cyclo(Phe-Phe)	$C_{18}H_{18}N_2O_2$	1.59	1.59	224.92	8	1.38	1.38	0.56
Cyclo(Trp-Tyr)	$C_{20}H_{19}N_3O_3$	1.05	1.05	255.26	10	0.90	0.90	-0.25
Boc-Phe-Val	$C_{19}H_{28}N_2O_5$	2.88		286.49	9	2.61		3.43
Boc-Phe-Tyr	$C_{23}H_{28}N_2O_6$	2.79	-0.02	324.82	10	3.11		3.68
AcTyrTyrNH$_2$	$C_{20}H_{23}N_3O_5$	-0.16	-0.16	284.86	11	1.12	1.12	-0.15
AcTyr(OMe)Tyr(OMe)NH$_2$	$C_{22}H_{27}N_3O_5$	1.19	1.19	313.96	11	2.04	2.04	1.15
Renin inhibitor 11	$C_{34}H_{50}N_8O_7$	2.85	2.85	507.73	18	3.15	3.15	1.92
Renin inhibitor 12	$C_{36}H_{51}N_7O_9$	2.11	2.11	530.44	21	1.71	1.71	0.37
Renin inhibitor 14	$C_{36}H_{50}N_8O_7$	2.20	2.20	513.89	20	1.90	1.90	1.62
Renin inhibitor 15	$C_{34}H_{50}N_8O_7$	1.94	1.94	507.54	20	1.70	1.70	0.57
Renin inhibitor 16	$C_{36}H_{53}N_7O_6$	1.42	1.42	542.29	23	0.64	0.64	-0.10
Renin inhibitor 17	$C_{36}H_{52}N_8O_7$	2.45	2.45	519.99	20	2.09	2.09	1.20
Renin inhibitor 18	$C_{36}H_{55}N_7O_9$	2.30	2.30	556.11	23	1.80	1.80	1.27
Renin inhibitor 19	$C_{36}H_{54}N_8O_9$	2.49	2.49	540.24	22	2.01	2.01	1.32
Renin inhibitor 20	$C_{38}H_{56}N_7O_{10}$	2.03	2.03	589.56	25	1.42	1.42	1.59
Renin inhibitor 21	$C_{38}H_{56}N_8O_9$	2.21	2.21	573.79	24	1.64	1.64	1.64
Renin inhibitor 22	$C_{39}H_{51}N_7O_6$	1.22	1.22	536.81	23	0.46	0.46	-0.93
Renin inhibitor 23	$C_{39}H_{56}N_7O_{11}$	2.06	2.06	606.08	27	1.22	1.22	0.27
Renin inhibitor 24	$C_{39}H_{58}N_8O_{10}$	2.31	2.31	587.05	26	1.33	1.33	2.26

Table 5.1—continued

Name	Formula	log *P*	log D^a	*v* (Å^3)	%	log P_{pr}	log D_{pr}[b]	log P_{ACD}
Renin inhibitor 25	$C_{43}H_{67}N_7O_{13}$	1.67	1.67	672.38	31	0.44	0.44	0.89
Renin inhibitor 26	$C_{43}H_{66}N_6O_{12}$	1.87	1.87	654.10	30	0.58	0.58	2.89
Renin inhibitor b23	$C_{38}H_{60}N_6O_7$	2.49	2.49	514.10	20	1.91	1.91	2.01
Renin inhibitor b24	$C_{38}H_{62}N_6O_7$	2.16	2.16	526.41	20	2.30	2.30	1.49
Renin inhibitor b25	$C_{38}H_{48}N_6O_7S_1$	2.83	2.83	521.68	20	2.15	2.15	2.14
Renin inhibitor b26	$C_{38}H_{60}N_4O_7S_1$	3.59	3.59	523.22	18	3.64	3.64	3.59
Renin inhibitor b27	$C_{38}H_{63}N_5O_6$	3.51	3.51	506.67	18	3.12	3.12	3.20
Renin inhibitor b28	$C_{38}H_{63}N_5O_6$	4.01	4.01	507.02	18	3.13	3.13	3.59
Renin inhibitor b29	$C_{38}H_{65}N_5O_6$	3.71	3.71	519.15	18	3.51	3.51	3.07
Renin inhibitor b30	$C_{38}H_{52}N_4O_6S_1$	4.27	4.27	514.43	18	3.36	3.36	3.72
Renin inhibitor b32	$C_{38}H_{63}N_5O_6$	3.79	3.79	505.69	18	3.08	3.08	3.76
Renin inhibitor b33	$C_{32}H_{50}N_6O_6$	4.10	4.10	484.61	16	3.86	3.86	4.16
Renin inhibitor c23	$C_{37}H_{54}N_6O_6S_1$	4.01	4.01	541.89	18	4.24	4.24	3.18
Renin inhibitor c25	$C_{38}H_{65}N_5O_5S_1$	4.22	4.22	555.51	18	4.67	4.67	3.99
AcValAlaAlaNH$_2$	$C_{13}H_{24}N_4O_4$	-1.40	-1.40	234.22	12	-1.22	-1.22	-2.09
AcValAlaValNH$_2$	$C_{15}H_{28}N_4O_4$	-0.67	-0.67	262.37	12	-0.32	-0.32	-1.21
AcValIleGlyNH$_2$	$C_{15}H_{28}N_4O_4$	-0.45	-0.45	262.03	12	-0.33	-0.33	-1.03
AcAlaLeuValNH$_2$	$C_{16}H_{30}N_4O_4$	-0.14	-0.14	275.91	12	0.11	0.11	-0.68
AcValPheAlaNH$_2$	$C_{16}H_{26}N_4O_4$	0.06	0.06	293.65	12	0.67	0.67	-0.16
AcAlaValIleNH$_2$	$C_{16}H_{30}N_4O_4$	-0.20	-0.20	275.90	12	0.11	0.11	-0.68
AcIlePheAlaNH$_2$	$C_{20}H_{30}N_4O_4$	0.52	0.52	307.67	12	1.12	1.12	0.37
AcGlyAlaValNH$_2$	$C_{12}H_{22}N_4O_4$	-1.56	-1.56	219.93	12	-1.67	-1.67	-2.44
AcAlaGlyPheNH$_2$	$C_{16}H_{22}N_4O_4$	-0.71	-0.71	251.08	12	-0.68	-0.68	-1.43
AcIleAlaValNH$_2$	$C_{16}H_{30}N_4O_4$	-0.21	-0.21	275.89	12	0.11	0.11	-0.68
AcPheGlyLeuNH$_2$	$C_{16}H_{26}N_4O_4$	0.60	0.60	293.80	12	0.68	0.68	0.03
AcPheIleGlyNH$_2$	$C_{16}H_{26}N_4O_4$	0.34	0.34	293.54	12	0.67	0.67	0.03
AcValValIleNH$_2$	$C_{16}H_{34}N_4O_4$	0.49	0.49	304.12	12	1.01	1.01	0.20
AcGlyLeuGlyNH$_2$	$C_{12}H_{22}N_4O_4$	-1.23	-1.23	219.80	12	-1.68	-1.68	-2.25
AcAlaTyrLeuNH$_2$	$C_{20}H_{30}N_4O_5$	-0.04	-0.04	313.65	13	0.59	0.59	-0.36
AcAlaTyrPheNH$_2$	$C_{23}H_{28}N_4O_5$	0.26	0.26	331.26	13	1.15	1.15	0.11
AcTrpAlaAlaNH$_2$	$C_{18}H_{25}N_5O_4$	-0.38	-0.38	289.07	13	-0.19	-0.19	-1.11
AcTrpIleGlyNH$_2$	$C_{21}H_{29}N_5O_4$	0.62	0.62	317.10	13	0.70	0.70	-0.05
AcTrpGlyPheNH$_2$	$C_{24}H_{27}N_5O_4$	0.99	0.99	334.81	13	1.26	1.26	0.42
AcTrpAlaValNH$_2$	$C_{21}H_{29}N_5O_4$	0.36	0.36	317.13	13	0.70	0.70	-0.23
AcAlaMetValNH$_2$	$C_{18}H_{28}N_4O_4S_1$	-0.63	-0.63	278.97	13	-0.52	-0.52	-1.04
AcIleMetPheNH$_2$	$C_{22}H_{34}N_4O_4S_1$	1.28	1.28	352.19	13	1.81	1.81	1.38
AcLeuSerPheNH$_2$	$C_{20}H_{30}N_4O_5$	0.23	0.23	314.24	13	0.61	0.61	0.01
AcLeuThrLeuNH$_2$	$C_{16}H_{30}N_4O_5$	0.24	0.24	311.07	13	0.51	0.51	-0.11
AcLysPheValNH$_2$	$C_{22}H_{36}N_5O_4$	0.51	-2.13	346.00	14	0.89	-1.27	-0.52
AcLysIlePheNH$_2$	$C_{23}H_{37}N_5O_4$	0.65	-1.46	360.16	14	1.34	-0.82	-0.03
AcLysPheLeuNH$_2$	$C_{23}H_{37}N_5O_4$	0.98	-1.51	360.22	14	1.35	-0.82	0.02
AcLeuLysPheNH$_2$	$C_{23}H_{37}N_5O_4$	1.14	-1.41	359.95	14	1.34	-0.83	-0.03
AcOrnIleValNH$_2$	$C_{19}H_{36}N_5O_4$	-0.43		314.62	14	-0.10	-2.27	-0.96
AcOrnPheLeuNH$_2$	$C_{22}H_{36}N_5O_4$	0.54	-1.37	345.49	14	0.88	-1.29	0.09
AcLeuOrnPheNH$_2$	$C_{22}H_{36}N_5O_4$	0.53	-1.38	345.56	14	0.88	-1.29	0.04
AcArgIlePheNH$_2$	$C_{23}H_{37}N_7O_4$	-0.04	-0.90	372.47	17	-0.43	-1.15	-0.77
AcArgPheLeuNH$_2$	$C_{23}H_{37}N_7O_4$	-0.03	-1.04	372.36	17	-0.43	-1.16	-0.72
AcLeuArgPheNH$_2$	$C_{23}H_{37}N_7O_4$	-0.06	-0.76	372.68	17	-0.42	-1.15	-0.77

Table 5.1—continued

Name	Formula	log *P*	log D^a	*v* (Å^3)	%	log P_{Pr}	log D_{Pr}[b]	log P_{ACD}
AcLeuPheArgNH$_2$	$C_{23}H_{37}N_7O_4$	-0.12	-0.93	373.15	17	-0.41	-1.13	-0.72
AcIlePheArgNH$_2$	$C_{23}H_{37}N_7O_4$	-0.05	-0.93	372.59	17	-0.43	-1.15	-0.72
AcHisIlePheNH$_2$	$C_{23}H_{32}N_6O_4$	0.44	0.36	348.80	14	0.98	0.98	-0.04
AcPheHisLeuNH$_2$	$C_{23}H_{32}N_6O_4$	0.64	0.46	348.89	14	0.99	0.99	0.01
AcIleHisValNH$_2$	$C_{16}H_{32}N_6O_4$	-0.24	-0.33	317.41	14	-0.02	-0.02	-1.05
AcGlyPheHisNH$_2$	$C_{19}H_{24}N_6O_4$	-0.92	-1.09	293.02	14	-0.79	-0.79	-1.80
AcTrpHisValNH$_2$	$C_{24}H_{31}N_7O_4$	0.31	0.16	358.17	15	0.56	0.56	-0.60
AcPheTrpHisNH$_2$	$C_{28}H_{31}N_7O_4$	1.00	0.89	390.24	15	1.58	1.58	0.40
AcAspPheLeuNH$_2$	$C_{21}H_{30}N_4O_6$	0.40	-1.39	329.15	14	0.36	-1.81	0.49
AcAspIlePheNH$_2$	$C_{21}H_{30}N_4O_6$	0.61	-1.32	329.51	14	0.37	-1.80	0.45
AcPheAspLeuNH$_2$	$C_{21}H_{30}N_4O_6$	0.39	-1.19	329.51	14	0.37	-1.80	0.49
AcLeuAspLeuNH$_2$	$C_{18}H_{32}N_4O_6$	0.16	-1.55	311.80	14	-0.19	-2.36	-0.03
AcIleLeuAspNH$_2$	$C_{18}H_{32}N_4O_6$	-0.04	-1.90	311.95	14	-0.19	-2.36	-0.03
AcGluPheLeuNH$_2$	$C_{22}H_{32}N_4O_6$	0.32	-1.52	343.53	14	0.82	-1.35	0.40
AcGluIlePheNH$_2$	$C_{22}H_{32}N_4O_6$	0.43	-1.57	343.29	14	0.81	-1.36	0.35
AcPheGluPheNH$_2$	$C_{25}H_{30}N_4O_6$	0.65	-1.08	361.04	14	1.37	-0.80	0.87
AcLeuGluPheNH$_2$	$C_{22}H_{32}N_4O_6$	0.48	-1.25	343.18	14	0.80	-1.36	0.35
AcLeuIleGluNH$_2$	$C_{18}H_{34}N_4O_6$	0.07	-1.87	325.47	14	0.24	-1.93	-0.12
AcAla-Gly-AlaNH*t*But	$C_{14}H_{26}N_4O_4$	-0.60	-0.60	248.03	12	-0.78	-0.78	-1.90
AcAla-Ala-AlaNH*t*But	$C_{15}H_{28}N_4O_4$	-0.51	-0.51	262.05	12	-0.33	-0.33	-1.56
AcAla-Phe-AlaNH*t*But	$C_{21}H_{32}N_4O_4$	1.01	1.01	321.86	12	1.57	1.57	0.38
AcAla-Trp-AlaNH*t*But	$C_{23}H_{33}N_5O_4$	1.25	1.25	344.98	13	1.58	1.58	0.30
AcAla-Pro-AlaNH*t*But	$C_{17}H_{30}N_4O_4$	-0.39	-0.39	281.89	13	-0.42	-0.42	-0.28
AcAla-His-AlaNH*t*But	$C_{18}H_{30}N_6O_4$	-0.48	-0.48	303.27	14	-0.47	-0.47	-1.92
AcAla-Asp-AlaNH*t*But	$C_{18}H_{30}N_4O_6$	-0.74	-3.40	283.65	14	-1.09	-3.26	-1.43
AcAla-Glu-AlaNH*t*But	$C_{17}H_{30}N_4O_6$	-0.67	-3.15	297.65	14	-0.64	-2.81	-1.53
AcPhe-Phe-PheNH$_2$	$C_{29}H_{32}N_4O_4$	2.30	2.30	383.53	12	3.53	3.53	2.77
AcFF(Me)FNH$_2$	$C_{30}H_{34}N_4O_4$	2.63	2.63	397.95	13	3.27	3.27	3.08
AcF(Me)F(Me)FNH$_2$	$C_{31}H_{36}N_4O_4$	2.53	2.53	412.86	14	3.02	3.02	3.38
Ac(Me)F(Me)F(Me)FNH$_2$	$C_{32}H_{38}N_4O_4$	2.92	2.92	426.90	15	2.75	2.75	3.69
Ac(Me)F(Me)F(Me)FNH(Me)	$C_{33}H_{40}N_4O_4$	3.24	3.24	440.99	15	3.19	3.19	3.92
TRH (pyroGln-His-ProNH$_2$)	$C_{16}H_{22}N_6O_4$	-2.46	-2.46	258.75	14	-1.88	-1.88	-3.09
N-Methyloxycarbonyl TRH	$C_{16}H_{24}N_6O_6$	-1.88	-1.88	294.95	15	-1.45	-1.45	-2.67
N-Ethyloxycarbonyl TRH	$C_{16}H_{26}N_6O_6$	-1.30	-1.30	309.29	15	-1.00	-1.00	-2.14
N-Isopropyloxycarbonyl TRH	$C_{20}H_{28}N_6O_6$	-0.80	-0.80	323.29	15	-0.55	-0.55	-1.79
N-Butyloxycarbonyl TRH	$C_{21}H_{30}N_6O_6$	-0.47	-0.47	336.74	15	-0.12	-0.12	-1.08
N-Isobutyloxycarbonyl TRH	$C_{21}H_{30}N_6O_6$	-0.44	-0.44	337.52	15	-0.10	-0.10	-1.26
N-Hexyloxycarbonyl TRH	$C_{23}H_{34}N_6O_6$	0.71	0.71	365.17	15	0.78	0.78	-0.01
N-Octyloxycarbonyl TRH	$C_{25}H_{38}N_6O_6$	1.88	1.88	392.82	15	1.66	1.66	1.05
N-(2-Et)Hexyloxycarbonyl TRH	$C_{25}H_{38}N_6O_6$	1.82	1.82	393.54	15	1.68	1.68	0.87
N-Cyclohexyloxycarbonyl TRH	$C_{23}H_{32}N_6O_6$	0.60	0.60	356.69	15	0.51	0.51	-0.61
N-Benzyloxycarbonyl TRH	$C_{24}H_{28}N_6O_6$	-0.11	-0.11	354.80	15	0.45	0.45	-0.89
IBHTrp-Met-Asp-PheNH$_2$	$C_{36}H_{43}I_1N_6O_5S_1$		1.60	572.72	20	3.77	1.60	4.37
IBHTrp-Met-Arg-PheNH$_2$	$C_{40}H_{50}I_1N_9O_4S_1$		1.90	618.83	23	3.07	0.90	3.15
IBHTrp-Asp-Met-PheNH$_2$	$C_{36}H_{43}I_1N_6O_5S_1$		1.70	576.29	20	3.89	1.72	4.37
IBHTrp-Asp-Nle-PheNH$_2$	$C_{36}H_{45}I_1N_6O_5$		1.30	574.62	19	4.56	2.39	4.91
IBHSer-Gln-Asp-GlyNH$_2$	$C_{23}H_{31}I_1N_6O_{10}$		-2.40	424.42	22	-2.39	-4.56	-2.00
IBHSer-Gln-Arg-GlyNH$_2$	$C_{25}H_{36}I_1N_9O_6$		-2.40	468.04	25	-3.17	-3.89	-3.22

Table 5.1—continued

Name	Formula	log *P*	log *D*[a]	*v* (Å^3)	π	log P_{pr}	log D_{pr}[b]	log P_{ACD}
BocPhe-His-Sta-Leu-Res9	$C_{36}H_{56}N_6O_7$		0.95	565.26	21	2.81	0.64	2.00
BocPhe-His-Sta-Leu-Res10	$C_{37}H_{58}N_{12}O_7$		0.32	591.67	24	1.49	0.76	1.68
BocPhe-His-Sta-Leu-Res11	$C_{40}H_{62}N_6O_7$		1.89	625.22	21	4.72	2.55	3.32
BocPhe-His-Sta-Leu-Res13	$C_{43}H_{64}N_{16}O_7$		1.02	652.78	24	3.43	2.71	3.02
AcTrpLeuAlaLeuLeu	$C_{34}H_{52}N_6O_7$	3.32		518.43	18	3.49		3.11
AcTrpLeuAspLeuLeu	$C_{38}H_{52}N_6O_8$	3.07		540.30	20	2.74		3.23
AcTrpLeuGluLeuLeu	$C_{39}H_{54}N_6O_8$	3.24		554.28	20	3.19		3.14
AcTrpLeuGlyLeuLeu	$C_{33}H_{50}N_6O_7$	3.26		504.42	18	3.04		2.77
AcTrpLeuSerLeuLeu	$C_{34}H_{52}N_6O_8$	3.10		525.17	19	2.98		2.80
AcTrpLeuThrLeuLeu	$C_{35}H_{54}N_6O_8$	3.27		537.55	19	3.38		3.15
(Me)Arg-Lys-Pro-Trp-*f*Leu-LeuOEt	$C_{43}H_{71}N_{11}O_7$	-0.05	-1.68	679.19	28	1.38		2.47
Cyclic Peptides								
RA-V (Rubiae Radix)	$C_{40}H_{48}N_4O_9$	2.72	2.72	561.18	24	0.52	0.52	3.62
RA-VII	$C_{41}H_{50}N_4O_9$	3.17	3.17	574.41	24	0.94	0.94	3.28
RA-V acetate	$C_{42}H_{50}N_4O_{10}$	3.19	3.19	590.50	25	0.73	0.73	3.56
Melanotan-II	$C_{50}H_{69}N_{15}O_9$	1.10	0.45	764.98	36	-1.67	-2.39	-7.45
Desmopressin (Mpa-Tyr-Phe-Gln-Asn-Cys-Pro-dArg-GlyNH_2)	$C_{46}H_{64}N_{14}O_{12}S_2$			767.79	40	-4.47	-5.19	-8.29
Sandostatin (Octreotide)	$C_{49}H_{66}N_{10}O_{10}S_2$	1.45	-0.49	751.35	31	1.51	-0.66	-3.91
Desmopressin, O-Butanoyl	$C_{50}H_{70}N_{14}O_{13}S_2$		-2.25	825.23	41	-3.36	-4.09	-7.15
Desmopressin, O-Hexanoyl	$C_{52}H_{74}N_{14}O_{13}S_2$		-1.25	853.06	41	-2.48	-3.20	-6.09
Desmopressin, O-Octanoyl	$C_{54}H_{78}N_{14}O_{13}S_2$		0.41	881.67	41	-1.57	-2.29	-5.02
Desmopressin, O-2-Ethylhexanoyl	$C_{54}H_{78}N_{14}O_{13}S_2$		0.09	881.35	41	-1.58	-2.30	-5.21
Desmopressin, O-Isobutyloxycarbonyl	$C_{51}H_{72}N_{14}O_{14}S_2$		-1.77	846.21	42	-3.42	-4.14	-6.87
Gramicidin S c(VOLFP)$_2$	$C_{60}H_{92}N_{12}O_{10}$		2.51	904.57	36	2.77	2.77	-4.42
Gramicidin S an1 c(VKLFP)$_2$	$C_{62}H_{96}N_{12}O_{10}$		2.45	932.51	36	3.66	3.66	-4.03
Gramicidin S an1 c(VKLLP)$_2$	$C_{56}H_{100}N_{12}O_{10}$		1.44	897.49	36	2.55	2.55	-4.60
Gramicidin S an1 c(VKLYP)$_2$	$C_{62}H_{96}N_{12}O_{12}$		1.37	930.07	38	2.14	2.14	-5.51
Gramicidin S an1 c(VKLAP)$_2$	$C_{50}H_{88}N_{12}O_{10}$		-0.49	813.64	36	-0.12	-0.12	-7.42
Gramicidin S an1 c(VKLSP)$_2$	$C_{50}H_{88}N_{12}O_{12}$		-1.16	827.03	38	-1.14	-1.14	-9.53
Gramic. S an1 c(VOLchAP)$_2$	$C_{60}H_{104}N_{12}O_{10}$		3.48	924.49	36	3.41	3.41	-2.62
Gramicidin S an2 c(LKLPG)$_2$	$C_{50}H_{88}N_{12}O_{10}$		-0.26	813.59	36	-0.12	-0.12	-7.34
Gramicidin S an2 c(LKLPA)$_2$	$C_{52}H_{92}N_{12}O_{10}$		0.33	842.17	36	0.79	0.79	-6.36
Gramicidin S an2 c(LKLPL)$_2$	$C_{56}H_{104}N_{12}O_{10}$		2.05	911.78	36	3.00	3.00	-3.54
Gramicidin S an2 c(LKLPF)$_2$	$C_{62}H_{100}N_{12}O_{10}$		2.32	947.84	36	4.15	4.15	-2.97
Gramic. S an2 c(LKLPchA)$_2$	$C_{64}H_{112}N_{12}O_{10}$		3.63	977.43	36	5.09	5.09	-1.17
Cyclosporin A	$C_{62}H_{111}N_{11}O_{12}$	2.96	2.96	971.32	41	1.28	1.28	11.28

[a] Log *D* values are distribution coefficients measured at neutral pH.
[b] Predicted distribution coefficients. Values are in bold if a correction for an ionizable side-chain was included.

Apparent Partition Coefficients (Monoionic Species)

When ionizable side chains are present, log *P* measurements and interpretations become much more complex (Taylor, 1990) as apparent partition coefficients become pH- and buffer-dependent. This often neglected aspect can lead to different hydrophobic contributions for these side chains. Ionization makes the apparent partition coefficient measured at neutral pH smaller than that of the neutral form, but several other issues must be considered as well. Two pH-dependence studies by Kim and Szoka (1992) for acetyl tripeptide *t*-butyl amides, respectively by Akamatsu et al. (1994) for acetyl di- and tripeptide amides indicate sigmoidal behavior; a lipophilicity profile also observed for prostaglandins (Avdeef et al., 1995). This behavior contradicts the usual assumption that only the uncharged form partitions in the octanol phase, and suggests that the ionized form – probably only as an ion-pair complex – also enters this phase (Yunger and Cramer, 1981; Taylor, 1990; Kim and Szoka, 1992; Akamatsu et al., 1994; Avdeef et al., 1995). At pH values sufficiently far from the compounds' pK_a the partition of the ionized form might even become the major contributor to the apparent partition coefficient. It is then not surprising that log *D* is also counterion, ionic strength and therefore buffer-dependent (Akamatsu et al., 1994). Consequently, different experimental results are inconsistent and often difficult to compare. We will briefly review the processes and equations that need to be considered.

Assuming that ionic species partition into the organic phase as ion-pairs, for ions A^- and B^+ the following equilibrium has to be described (Shim et al., 1981; Taylor, 1990):

$$A_w^- + B_w^+ \xleftrightarrow{K_p} (A^-B^+)_w \xleftrightarrow{P_p} (A^-B^+)_o \qquad (5.8)$$

Here K_p represents the ion pair association constant in the aqueous phase

$$K_p = \frac{[(A^-B^+)]_w}{[A^-]_w[B^+]_w} \qquad (5.9)$$

while P_p is the (true) partition coefficient of the ion pair

$$P_p = \frac{[(A^-B^+)]_o}{[(A^-B^+)]_w} \qquad (5.10)$$

Their product is the extraction constant

$$K_e = \frac{[(A^-B^+)]_o}{[A^-]_w[B^+]_w} = K_p P_p \qquad (5.11)$$

Introducing the simplifying assumption that only ion pairs exist in the organic phase, the molar concentration ratio of a species between two phases at equilibrium can be characterized here with an empirical extraction coefficient E_c

$$E_c = \frac{[A^{tot}]_o}{[A^{tot}]_w} = \frac{[(A^-B^+)]_o}{[A^-]_w + [(A^-B^+)]_w} \qquad (5.12)$$

This dimensionless quantity represent the analog of P defined in eq. 5.1 for a neutral species, but it depends on the nature and concentration of the

counterion. Assuming that ion pair formation in the aqueous phase is not significant, K_p is small, and we have

$$E_c = K_p P_p [B^+]_w \qquad (5.13)$$

For ionizable peptides partition of the neutral form and extraction of the ion pair has to be considered together. In other words, the measured distribution coefficient results from the distribution of the neutral form and the ion pair; depending on pH and counterion concentration, one or the other is dominant. This results in a sigmoidal lipophilicity profile that agrees well with those observed experimentally. The following additional assumptions are considered here (written for a monoacidic compound A):

- the equilibrium partition can be characterized with P (eq. 5.1) for the neutral form (A) respectively with E_c (eq. 5.12) for the ionic form (A^-)
- the ionic equilibrium in the aqueous phase can be characterized with the usual equilibrium constant

$$K_a = \frac{[A^-]_w [H^+]_w}{[A]_w} \qquad (5.14)$$

- the apparent partition coefficient or distribution coefficient (D) measured at some fixed condition is that of the neutral and ionic forms considered together (the technique of measurement cannot distinguish between the two forms):

$$D = \frac{[A]_o + [(A^- B^+)]_o}{[A]_w + [A^-]_w} \qquad (5.15)$$

As previously, the concentration of the ionic form in octanol and the concentration of the ion pair in the aqueous phase was neglected. Using equations 5.1, 5.12, and 5.14 to eliminate the corresponding concentrations one get:

$$D = \frac{P + E_c \frac{K_a}{[H^+]}}{1 + \frac{K_a}{[H^+]}} \qquad (5.16)$$

Figure 6.2 shows log D versus pH according to this equation that is similar to those used in other partition studies (Austin et al., 1995) of ionizing molecules and gives a sigmoidal profile which agrees well with those obtained experimentally (Kim and Szoka, 1992; Akamatsu et al., 1994; Avdeef et al., 1995). At pH values more than 3-4 log units away from the compounds' own pK_a the ion pair partition becomes dominant. On the other hand, if ion pair extraction is inappreciable, the ionic form cannot partition in the octanol phase ($E_c = 0$), and equation 5.16 reduces to the commonly used (Grant and Higuchi, 1990)

$$D = \frac{P}{1 + \frac{K_a}{[H^+]}} \qquad (5.17)$$

that is,

$$\log D = \log P - \log(1 + 10^{pH - pK_a}) \qquad (5.18)$$

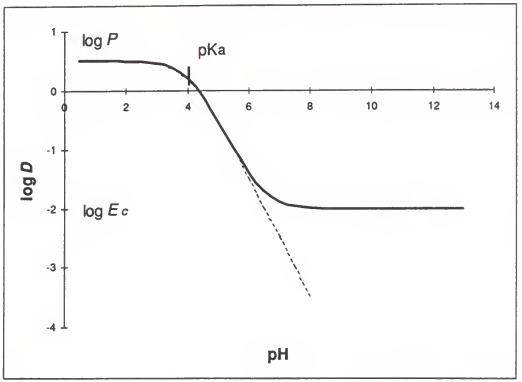


Figure 5.2 Apparent partition coefficient as a function of pH according to equation 16 for a hypothetical compound ($pK_a = 4$, $\log P = 0.5$, $\log E_c = -2.0$).The dashed line corresponds to the assumption frequently used for ion correction that ion pair extraction is completely negligible (eq. 5.17, 5.18).

This is the formula frequently used for ion correction, and it is shown by the dashed line in figure 5.2. For basic compounds all conclusions are similar, but with inverted pH dependence. Since NaCl is the dominant physiological electrolyte at a concentration not far from 0.1 mol/l, this could be a reasonable choice for standard conditions to compare partition properties of ionic compounds (Taylor, 1990). However, as no standard conditions are yet established and even physiological conditions are approximated in different ways, here we had to disregard the buffer dependence. As usual buffer concentrations are somewhere in the 0.01-0.1 M range, the introduced

inaccuracies might be even larger than one log unit, but most values were determined sufficiently close to a desired standard (low solute concentration, 0.1 M of low molecular weight counterion) to allow some comparison.

For aspartate, glutamate and histidine residues detailed studies were made in both reference 9 and 10. In agreement with the pK_a values of these amino acids, most of the change in the apparent partition coefficients occurs between pH 3 and 7. The log P - log E_c differences are relatively stable:

Asp (pK_a ~4): 1.77(±0.14) between pH 2 and 7 (Akamatsu et al., 1994);
respectively 2.65 between pH 2 and 10 (Kim and Szoka, 1992)

Glu (pK_a ~4): 1.85(±0.12) between pH 2 and 7 (Akamatsu et al., 1994);
respectively 2.48 between pH 2 and 10 (Kim and Szoka, 1992)

His (pK_a ~6.5): ~1.30 between pH 12 and 4 (Akamatsu et al., 1994); respectively 1.57 between pH 9 and 3 (Kim and Szoka, 1992)

These differences are not too large; certainly smaller than those observed between the partition of the neutral and the ionic form of amine or acid functions in other molecules (Leo et al., 1971; Rekker and de Kort, 1979). The special flexibility and hydrogen-bonding ability of peptides might play a role in this. Ionization of the acid terminal in acetylpentapeptides seems to cause a higher difference. The average shift between pH 1 and 9, estimated from published (Wimley et al., 1996) mole-fraction free energies of transfer (log P = log (v_w/v_o) + $\Delta G/(2.303RT)$), is around 3.50, and is quite close to that around 4 observed for usual organic -COOH fragments (Leo et al., 1971; Rekker and de Kort, 1979).

However, the additional difference due to a second charge on an aspartate or glutamate side chain in these acetyl-pentapeptides is again about the same as mentioned earlier: 2.31, respectively 2.54.

The side chain of histidine residues is prevalently in neutral form at neutral pH, thus, the apparent partition coefficient measured at this pH is very close to that of the neutral form – the average observed difference is only 0.13(±0.04) (Akamatsu et al., 1994). Therefore, if only log *D* at pH=7 or 7.4 was available for this side chain, we considered it as log *P* of the neutral form.

Unfortunately less data and from only one study are available for the other basic residues (arginine, lysine, and ornithine). The log *P* - log E_c differences obtained from Akamatsu et al. (1994) are:

Arg (pK_a ~13): 0.85(±0.11), between pH 13 (ionization corrected) and 7

Lys (pK_a ~10.5): 2.45(±0.19), between pH 13 and 7

Orn (pK_a ~10.5): 1.88(±0.05), between pH 13 and 7

All these values are no doubt influenced not only by the buffer used, but also by the nature of neighboring side chains, by conformation, and by other factors as well. Solvation of ionic species may also involve interactions not adequately described by a model developed for neutral species, but in a first approximation within the present model these shifts can be characterized by using a change in π contributions. When computing an estimate for log *D* at pH = 7, we will use a $\Delta\pi$ = -3 correction for acidic side chains (Asp, Glu) causing a shift of -2.18 compared to the neutral form. We will use a change of $\Delta\pi$ = -1

(-0.73) for arginine, and $\Delta\pi$ = -3 (-2.18) for lysine and ornithine. These are only rough, first estimates to describe the octanol-water distribution of charged species. All conclusions are for nonzwitterionic peptides only, that is, peptides that do not contain ionized basic and acidic parts simultaneously. There are not yet enough consistent partition data for zwitterionic, respectively multiply charged compounds to establish a clear relationship; especially that spatial conformation, intramolecular bonding, and cosolvation have more significant roles here. For two similar charges on the same molecule at relatively distant places additive contributions might represent a reasonable choice, but additivity certainly does not hold when opposite charges are present.

For the total set of 190 apparent partition coefficients determined around neutral pH for un-ionizable (150) or monoionic (40) peptides we obtained this way a reasonable correlation (table 5.1):

$$\log D = -0.114(\pm 0.039) + 0.925(\pm 0.023) \log D_{pred}$$
$$n = 190;\ r^2 = 0.899;\ \sigma = 0.533;\ F = 1681.3 \qquad (5.19)$$

Cyclic Peptides

Solubility properties are strongly influenced by the existence of intramolecular hydrogen bonds. In the aqueous phase, where the surrounding water molecules provide many donor and acceptor sites, intramolecular hydrogen bonding is more favorable for peptides that already have a cyclic structure. Here, the already existing rigidity of the structure makes the entropy loss due to additional transannular H bond formation less unfavorable. If these

bonds are maintained in the aqueous phase, water solubility will decrease, and consequently hydrophobicity and log *P* will increase. Cyclization of peptides was long known as a method to obtain more lipophilic forms. A predictive model that does not account for such hydrogen bonds is likely to underestimate the real partition coefficient. It is almost impossible to predict the number of such bonds in larger molecules; even conformational studies can be inconclusive. We present here a few large, cyclic peptides with available experimental values comparing the measured and calculated partition data in the light of those discussed above. Most differences between measured and calculated values are in reasonable agreement with the number of transannular hydrogen bonds determined by other conformational studies. This is additional evidence for the relation between the π parameter used in this model and the number of hydrogen bonds formed during the octanol-water transfer.

Cyclosporin A

The present model predicts a partition coefficient of 1.28 for this immunosuppressive cyclic undecapeptide of fungal origin that is far larger than any molecule included in the training set. Its log *P* has been estimated (Cefalu and Pardridge, 1985; El Tayar et al., 1993) to be 2.92-2.99. However, an HPLC determination (Takada et al., 1992) suggests log *P* = 0.55. This difference, respectively that between the calculated and the measured value, is quite close to that resulting from the possibility of three internal H-bonds (El Tayar et al., 1993) – causing an increase around two units in log *P* according to most

methods (three times the usual H-bonding correction of 0.6-0.7). The present prediction is by all accounts orders of magnitude better than the unrealistic values of CLOGP (14.0) or ACD/LogP (11.3).

Antitumor Hexapeptides RA

Antitumor hexapeptides RA are cyclic, unionizable hexapeptides obtained (Itokawa et al., 1984) from *Rubiae Radix* and *Rubiaceae*. The present model underestimates log *P* for the three derivatives where they were experimentally determined in average by 2.30 (RA-V 2.72/0.52, RA-VII 3.17/0.94, RA-V acetate 3.19/0.73). This difference is quite large, but conformational analysis (Morita et al., 1991) for this series indicate two intramolecular H bonds in crystalline structure and in solution as well, leaving unaccounted a difference of only about one unit. It is also worth noting that these experimental values were determined by HPLC essays, and such results need to be treated with precautions for such large molecules, as the case of cyclosporin proves it. For this reason only values obtained with usual shake-flask measurements were included in the correlations.

Gramicidin S Analogs

Experimental data for 12 analogs of gramicidin S, a symmetric, cyclic decapeptide with a repeated pentapeptide sequence, were available (Katayama et al., 1994). All structures contain two lysine or two ornithine residues, and only the apparent partition coefficient could be determined, as log *P* for the neutral form was too high for reliable measurements. These peptides have an antiparallel β-pleated sheet structure stabilized by two pairs of transannular

hydrogen bonds. In addition, each NH_3^+ group of two Orn residues can hydrogen-bond with the backbone carbonyl of the D-Phe residues (Krauss and Chan, 1982; Katayama et al., 1994). Assuming that when computing the apparent partition coefficients the effects of the two positively charged Lys or Orn residues can be described additively with the $\Delta\pi = -3$ correction established earlier, and that $\Delta\pi = +6$ for the mentioned six transannular hydrogen bonds, the calculated values are in very good agreement with the measured ones: $n = 12$, $r^2 = 0.922$, $\sigma = 0.458$, $F = 117.8$. As these compounds are much larger than those used in developing the model, this agreement can be considered a good confirmation of the previous assumptions. The predictions of ACD/LogP are in this case six or seven orders of magnitude lower (table 5.1).

Desmopressin Derivatives

Desmopressin (1-desamino-8-D arginine vasopressin, DDAVP) is a partially cyclic octapeptide with high antidiuretic activity and contains a hexapeptide ring structure held together by a disulfide bridge. Using the correction described earlier for charged arginine residues, its apparent log *D* is underestimated with about 2.2 units for the five ester derivatives (Kahns et al., 1993) considered. Again, the presence of transannular hydrogen bonds in these structures might account for most of this difference. A recent conformation study (Kihlberg et al., 1995) in aqueous solution indicates a lowest energy conformation with at least two hydrogen bonds between Phe^3 CO and Asn^5 NH, and between Gln^4 CO and Cys^6 NH. The crystal structure of pressinoic acid

seems to have three internal H bonds with differently positioned β-turns (Langs et al., 1986).

Melanotan-II

Melanotan-II is a cyclic heptapeptide analogue (Lan et al., 1994) of α-melanocyte-stimulating hormone that tans the skin. It contains two basic residues (Arg, His) and a partial ring structure held together by a lactam bridge of the side chain of the aspartate and lysine residues. Its partition coefficient is underestimated with 2.8 units by the present method, but intramolecular hydrogen bonds might again be responsible for part of the difference. The prediction of ACD/LogP is much lower, giving an error of 8.6 log units. The apparent partition coefficients measured here seem to reinforce the additivity assumed for similar charges and the shifts assumed earlier for arginine and histidine. The decrease in log *D* between pH 12.2 and 7.35 is -0.65 (close to ~-0.8 of arginine and 0 for histidine), while the decrease between pH 7.35 and 4.57 is -1.39 (again close to the presumed ~-1.5 for histidine in this pH range).

Sandostatin (Octreotide)

For the somatostatin analogue octapeptide sandostatin, the partition coefficient determined here for the neutral form agrees well with the predicted one (1.45 vs 1.64). The value obtained at pH 7.0 and 0.1 M NaCl (-0.49) also agrees with the distribution coefficient (-0.56) predicted using the average shift described earlier for a lysine residue. The measured distribution coefficient is relatively close to the only available literature data (Drewe et al., 1993) obtained

at a different temperature and with a different buffer: -0.22 (pH 7.5, 0.3 M mannitol + 20 mM HEPES-Tris buffer, 37°C). The ACD/LogP predicted value is again orders of magnitude lower (-3.91).

There seems to be enough evidence suggesting that the fully computerized and distinctively simple (two-parameter) model developed for smaller organic molecules using a molecular size-based approach can be useful in estimating partition coefficients for neutral peptides, even for quite large ones. An extension of the model also seems possible to estimate apparent partition coefficients of nonzwitterionic peptides at neutral pH, but additional precaution is suggested. Since partition here is counterion-dependent, and as no standards are yet defined, many effects were neglected and some assumptions are based on only a small number of experimental data. Predictions for larger, cyclic peptides cannot be considered very reliable, but they seem orders of magnitude more realistic than those obtained from two-dimensional fragment methods, and in many cases most of the error might be caused by intramolecular hydrogen bonding. Part of the result confirmed again the assumption that the π parameter introduced in the model is related to the change produced in hydrogen bonding by the octanol-water transport.

CHAPTER 6
RETROMETABOLIC DRUG DESIGN: PHYSICOCHEMICAL CONSIDERATIONS

Background

Physicochemical considerations, just as metabolism and targeting considerations, should be an integral part of any drug design process right from the beginning. In the fascination of designing three-dimensional models that fit some receptor, or at least the idea of what that might be, medicinal chemists tend to forget that before getting even close to its designated target, a molecule has to face many hurdles: it not only has to by-pass cells, membranes, and metabolizing enzymes, as mentioned earlier, but it also has to withstand solubility, stability, partition, distribution, and other challenges. A rigorous physicochemical approach might give much less exciting answers or design ideas, but, except for a limited number of cases where design of very effective agents is possible because pharmacological responses are produced via immense amplification (and thus, they can be triggered even by minuscule drug quantities reaching the receptors), basic physicochemical ideas might be very useful, as they can be applied for a much broader spectrum of problems. It is of no wonder, therefore, that despite a very late start, somewhere in the second half of this century, partition coefficients—especially between *n*-octanol and water (log $P_{o/w}$)—arose among the most successful parameters used in medicinal

chemistry. Even unglamorous molecular parameters, such as molecular volume, surface area, dipole moment, and polarizability, can reveal a great deal to an experienced eye. These parameters, which are today easily computable, considered together with hydrogen bonding ability and other relevant properties, could significantly reduce the number of failures faced by exclusively biological or chemical approaches when going from drawing boards or *in vitro* testing to *in vivo* applications.

In addition, experimental determination of the octanol-water partition coefficient can be considered now as part of the standard physicochemical characterization of novel compounds. Therefore, it is almost a requirement for a drug design or medicinal chemistry laboratory to have a tested and reliable experimental method for octanol-water partition and molecular lipophilicity measurements.

Our workgroup had a special interest in determining the lipophilicity of different chemical delivery systems (CDSs): their change during the NADH $\leftrightarrow$ NAD^+ type conversion of the 1,4-dihydrotrigonelline targetor moiety used in different brain targeting CDSs (Bodor and Brewster, 1983; Bodor et al., 1992; Bodor, 1995a; Bodor and Buchwald, 1997b), respectively their value for novel ketoxime analogues of β-adrenergic blockers that serve as their site- and stereo-specific CDS (Bodor and Prokai, 1990; Prokai et al., 1991; Bodor, 1995b). The computerized model, described in earlier chapters, proved useful here, and these developments will allow future designs to incorporate calculated

physicochemical properties even before actual synthesis is started. This chapter describes some of the basic principles for chemical delivery system (CDS) design and presents the pertaining log *P* calculations and experimental determinations.

Experimental Determinations

Chemicals and Solvents. The preparation of oximes has been described previously (Bodor et al., 1988, 1997a; Simay et al., 1989; Farag et al., 1997). Alprenolol was obtained from Sigma Chemicals Co. (St. Louis, MO), timolol maleate from Interchem Corp. (Paramus, NJ), and betaxolol from Alcon Laboratories Inc (Fort Worth, TX). Solvents used were of HPLC grade and supplied by Fisher Scientific (Fair Lawn, NJ, USA); additional chemicals used were commercially available, analytical reagent-grade products. Water was purified by reverse osmosis and ion exchange.

High-Performance Liquid Chromatography. The HPLC system consisted from a Spectra Physics (San Jose, CA, USA) SP8810 precision isocratic pump, SP8450 variable-wavelength UV-Vis detector, SP4290 computing integrator, and SP4290 autosampler. Separation was done on a 7.5 cm Supelcosil LC-8-DB (octyldimethylsilyl bonded phase) analytical column with 3-μm particle-size packing, coupled to a 2-cm guard column. All columns and accessories were supplied by Supelco (Bellofonte, PA, USA). Mobile phases were 20-40% acetonitrile in 5 mM ammonium acetate buffer (pH set to 4.0 with acetic acid). Flow rate was 0.6 ml/min. with an injection loop of 20 μl; detection was made at

254 nm. All measurements were made in triplicate. For peptide-CDSs best peak shapes and separations were obtained under the following conditions: The column used was a 15-cm Altima C_{18} analytical column with 5-μm particle-size packing. The mobile phase was 70% methanol, 15% chloroform, and 15% of the above buffer. Flow rate was 2 ml/min. with an injection loop of 20 μl and detection at 275 nm. The only disadvantage is that the mobile phase used is almost saturated with water and therefore sometimes the aqueous phase could not be reliably analyzed as injection of water perturbed the flow.

Partition Coefficient. Octanol-water distribution and partition coefficients were determined using the shake flask method according to standard requirements and recommendations (Taylor, 1990). As all compounds contained an ionizable amine function, a buffer was used in the aqueous phase to ensure stable pH. Here, commercially available pH 7.00 potassium phosphate monobasic sodium hydroxide buffer, respectively pH 10.00 potassium carbonate – potassium borate – potassium hydroxide buffer (FisherChemical, 0.05 M) were used. Both solvents (octanol and aqueous) were presaturated with the other by leaving them in contact for at least 24 hours. A small amount of solute was added to one of the solvents; after complete solvation it was filtered if necessary. Initial concentrations were as low as possible, around 0.1 mM. A known ratio of the other solvent was then added, and the mixture was vigorously shaken for long enough to reach equilibrium (more than 200 inversions, more than 10 minutes). After equilibration for a few minutes, the mixture was centrifuged (10 minutes at

2,000 rpm) to achieve good separation. Final concentrations in the initial and whenever possible in both phases were determined by absorption measurements using both UV spectrometry and HPLC procedures. The distribution coefficient was computed from the concentration ratios estimated from the absorption measurements correcting for unequal solvent volumes when they were used, e.g., if aqueous concentrations were measured

$$D = (c_{in} - c_{fin})/c_{fin} \times V_w/V_o \qquad (6.1)$$

The volume ratios were selected so to be close to equidistribution of compound in the two phases, and for each determination at least two different ratios were used. For ionizable species, assuming that only the neutral form partitions into the lipid (octanol) phase, the expected relationship between the log distribution coefficient (log *D*), measured at a given pH, and log *P* of the neutral form is the usual pK_a based ion correction (for bases):

$$\log P = \log D + \log (1 + 10^{pK_a - pH}) \qquad (6.2)$$

Concentration Determinations. Values in the study are from concentration measurements based on comparison of peak areas obtained with the described HPLC procedure. This method allows isomer separation, is accurate, and is reproducible if autosampler injection is used. For verification, the UV spectrum between 200 and 400 nm for 3 ml of solution was also recorded with a Perkin Elmer Lambda 11 UV/VIS spectrometer connected to a computer running the PECCS software package. For this method, concentrations have been evaluated

at different wavelengths around the absorption maximum above 250 nm and the obtained average was used.

Principles of Retrometabolic Drug Design

This section will briefly review some advanced chemical-enzymatic-based drug targeting systems obtained with strategies that are part of an approach designated now as "retrometabolic drug design" (RMDD). Data accumulated on metabolic activation-toxication and/or deactivation of biologically active chemicals show the necessity to include metabolic considerations in the drug design process. Instead of simply waiting for drug metabolism studies after selection of the best drug candidate, structure-activity relationships (SAR) should be combined with structure-metabolism relationships (SMR) right from the beginning and throughout the drug design process. This combination of structure-activity (SAR) and structure-metabolism relationships (SMR) is incorporated into the retrometabolic drug design concept. The RMDD loop includes *soft drug* (SD) design and *chemical delivery system* (CDS) design as shown in Fig. 6.1.

RMDD approaches thus include two distinct methods to improve the therapeutic index of a drug. One is the general concept of CDSs. A CDS is defined as a biologically inert molecule that requires several steps in its conversion to the active drug and that enhances drug delivery to a particular organ or site. The other side of the loop in Fig. 6.1 represents *soft drug* (SD) design strategies. Although both approaches involve chemical modifications to

obtain an improved therapeutic index and require enzymatic reactions to fulfill drug targeting, there is not much in common between the principles of soft drug design and those of CDS-based RMDD. While the CDS is inactive by definition and sequential enzymatic reactions provide the differential distribution and drug activation, soft drugs are active isosteric-isoelectronic analogs of the lead drug that are deactivated in a predictable and controllable way after they achieve their therapeutic role. In optimally extreme cases, for a CDS, enzymatic processes produce the drug at the site and nowhere else in the body, while in the case of the SD the drug is present at the site, but absent in the rest of the body due to enzymatic activity in those sites.

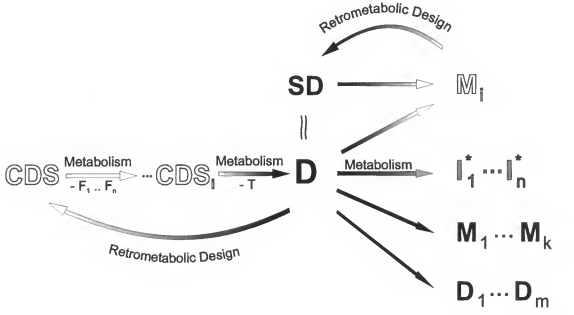


Figure 6.1 The retrometabolic drug design loop, including chemical delivery system (CDS) and soft drug (SD) design. Straight arrows represent metabolic changes; curved arrows represent retrometabolic design approaches. D drug, M_i inactive metabolites, $D_1..D_m$ analogue metabolites, $M_1..M_k$ other metabolites, $I^*_1..I^*_n$ potential reactive intermediates; CDS_i metabolites of the original CDS formed during the sequential metabolic conversion that removes the modifier functions (F) and the targetor moiety (T).

Chemical Delivery Systems

In principle, chemical drug delivery systems should include any drug targeting system that requires a chemical reaction to produce it. Consequently, they should include those systems where there is a covalent link between the drug and the so-called "carrier," and, accordingly, at least one chemical bond needs to be broken to release the active component. However, in the strictest sense, chemical drug delivery systems refer to inactive chemical derivatives of a drug where one or more chemical modifications were done, the newly attached moieties are monomolecular units (generally comparable in size to the derivatized molecule), and these modifications provide a site-specific or site-enhanced delivery of the drug by design. This requires multistep enzymatic and/or chemical transformations (Bodor, 1987; Bodor and Prokai, 1990). Some of these delivery systems have been reviewed before (Bodor and Brewster, 1983; Bodor and Kaminski, 1987; Bodor and Buchwald, 1997b).

During the chemical manipulations, two types of bioremovable moieties are introduced to convert the drug into an inactive precursor form: (1) a targetor (T) responsible for targeting, site-specificity, and lock-in; (2) modifier functions ($F_1...F_n$) that serve as lipophilizers, protect certain functions, or fine-tune the necessary molecular properties to prevent premature, unwanted metabolic conversions. Targeting is achieved by design: the CDS will undergo sequential metabolic conversions, disengaging the F and finally the T, after this moiety fulfills its site- or organ-targeting role. The CDS concept evolved from the

prodrug concept (Bundgaard, 1985; Bodor and Kaminski, 1987), but is essentially different. Within the present formalism, one can say that prodrugs contain one or more F moieties for protected or enhanced overall delivery, but they do not contain targetor moieties. Thus, they generally fail to achieve drug targeting, which is the major pathway to improve the therapeutic index. It is important to note again that the starting CDS and the intermediates leading to the active drug are all inactive and thus, do not produce any unwanted side effects. CDSs will be activated to the desired drug by a designed metabolic conversion at the site of action, and no other metabolism should take place.

With a CDS, targeting is achieved by design: recognizing specific enzymes found primarily, exclusively, or at higher activity at the site of action, or exploiting site-specific transport properties [e.g., blood-brain barrier (BBB)]. The strategically predicted, multienzymatic transformations result in a differential distribution of the drug. The concept of CDSs already has been applied in a variety of drug targeting problems, achieving successful delivery to the brain, to the eye, and to other organs (Bodor, 1994, Bodor and Buchwald, 1997) . CDSs can be divided into several classes as follows:

1. *Site-specific enzyme-activated CDSs* that exploit specific enzymes found primarily, exclusively, or at higher activity at the site of action.
2. *Enzymatic physical-chemical–based CDSs* that exploit site-specific traffic properties by sequential metabolic conversions resulting in altered transport properties.

3. *Receptor-based (transient anchor-type) CDSs* that enhance selectivity and activity through transient, reversible binding to the receptor.

Site-Specific Enzyme-Activated Chemical Delivery System

Design Principles

This type of CDS is based on enzymatic conversions of a CDS that take place only at the site of action as a result of differential distribution of certain enzymes within the body. If some specific enzymes are present only around the designated site of action, their use in a strategically designed system can lead to high site specificity. The drug is chemically modified, obviously also involving the pharmacophore, to introduce the targetor moiety and the eventual protective function(s). Following administration, the resulting CDS is distributed throughout the body. While it is continuously eliminated without producing any pharmacological effect from the "rest of the body," predictable enzymatic reactions convert the original CDS to the active drug only at the targeted site where the necessary enzymes are found primarily, exclusively, or at higher activity.

Certainly, the system depends upon the existence of such enzymes. The basis of the retrometabolic design that resulted in a successful site and stereospecific delivery of intraocular pressure (IOP)-reducing β-adrenergic blocking agents to the eye is a general metabolic process, which appears to apply to all lipophilic ketone precursors of β-amino alcohols. Noticing that the lipophilic esters of adrenalone, but not the free catechol-containing adrenalone,

are effectively reduced within the eye to yield epinephrine via the corresponding esters, whereas oral administration does not result in any formation of epinephrine, the ketone function was recognized as a targetor that can be reduced to produce the β-amino alcohol pharmacophore necessary for action. Since the approach could be extended to other β-agonists (Reddy and Bodor, 1994), the possibility arose to produce ophthalmically useful β-adrenergic antagonists within the eye from the corresponding ketones. However, ketones of the phenol ether type β-blockers are not good bioprecursors to produce the β-amino alcohols in these compounds because in aqueous solutions, they decompose to form the corresponding phenols. To stabilize them, they were converted to oximes, which need to undergo a facile enzymatic hydrolysis to the bioreducible ketone. Although the oximes are much more stable than the ketones, their aqueous stability still does not provide an acceptable shelf life (e.g., even at pH ~4.5, where the oxime is most stable, t_{90} is only about 2.5 months for alprenoxime). Additional and significant stabilization can be obtained, however, by using methoximes, methylethers of the corresponding oximes. A variety of such oxime and methoxime analogues of alprenolol, betaxolol, propranolol, timolol, etc., were synthesized and studied (Bodor et al., 1988, 1997a; Simay et al., 1989; Prokai et al., 1995, Farag et al., 1997). They were all shown to undergo the predicted specific activation within, and *only* within, the eye. Highest concentrations of the active β-antagonists were observed in the iris-ciliary body. When applied topically in rabbits, both the oximes and the

methoximes showed significantly higher IOP-reducing activities than the corresponding alcohols. On the other hand, topical or even i.v. administration did not produce any cardiovascular effect neither in normal rabbits, rats, and dogs, nor after inducing tachycardia. Following i.v. injection, the oxime disappeared rapidly from the blood, and at no time could the corresponding alcohol be detected (Bodor et al., 1988). This indicates that the required enzymatic hydrolysis–reduction activation sequence, which occurs in the eye, does not take place in the systemic circulation.

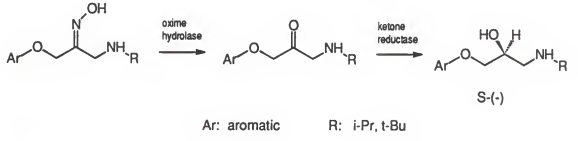


Figure 6.2 Sequential activation of oximes in the site- and stereospecific delivery of β-adrenergic antagonists to the eye. The original oximes and the intermediate ketones are inactive; they are enzymatically converted into the active S-(–) β-adrenergic blocker alcohols in a site- and stereospecific manner.

To summarize, in these CDSs, a β-amino oxime or alkyloxime function replaces the corresponding β-amino alcohol pharmacophore part of the original molecules (Fig. 6.2). This oxime or alkyloxime derivative is enzymatically hydrolyzed within the eye by enzymes located in the iris-ciliary body, and subsequently, reductive enzymes also located in the iris-ciliary body produce only the active S-(–) stereoisomer of the β-blocker (Bodor and Prokai, 1990). All these derivatives showed significant IOP-lowering activity, but even their i.v.

administration did not produce the active β-blockers metabolically. Therefore, they are of interest as potential antiglaucoma agents and are void of any cardiovascular activity, a major drawback of classical antiglaucoma agents. Methoxime derivatives provide sufficient stabilization for acceptable shelf life.

Physicochemical Data—Calculations and Measurements

Here we report an investigation of their lipophilicity based on octanol-water partition coefficients. The majority of data is for oxime analogs, but, for comparison, some corresponding alcohols and methoximes were also included. These oximes are known to exist in alternative Z (*syn*) or E (*anti*) configuration (Figure 6.3). The present study confirms that the orientation of the oxime hydroxyl or methoxyl group influences partition properties and pK_a values as well.

Reversed-phase high-performance liquid chromatography is a suitable technique for separation of E and Z isomers of different aryloxyalkylaminoketone-oximes and -methoximes, with the E isomers eluting first (Prokai et al., 1991). A special problem is introduced here by the fact that the oximes are present as two different (E and Z) isomers, and depending on compounds, buffers, and other conditions, the equilibrium isomer ratios are different (Bodor et al, 1997a; Farag et al., 1997). As the HPLC procedure allows isomer separation and the rate of isomerization in the simple organic buffers used is not very fast (re-establishment of the equilibrium isomer ratio requires several hours), the log octanol-water distribution and partition coefficients (log *D*

pH 7.0, log *P*) could be evaluated separately for the E and Z isomers. For example, the apparent rate of change (k_{obs}) of the concentration of the minor isomer of betaxolone oxime in buffer solution is 0.039 h^{-1} at pH 7.4 and 0.0275 h^{-1} at pH 6.5, respectively (Farag et al., 1997). It also worth noting that among the compounds studied timolone oxime and methoxime are unique; they do not show any isomerization whether in buffer, pH 6.5-7.4, or in biological fluids such as bovine aqueous humor, rat plasma, and rat liver homogenate (Bodor et al., 1997a).

Since the pK_a of all parent β-blockers (aryloxy-alkylaminoalcohols) is bellow 10 (alprenolol, betaxolol, propranolol: 9.50; metoprolol, timolol: 9.20) and that of the oximes is expected to be even lower, the distribution coefficient measured at pH 10.0 can be considered the partition coefficient of the neutral species (log *P*). This assumption was nicely confirmed by the excellent agreement between the log *P* values obtained here this way for the parent alcohols included and the published data; measured value/average of published values are: alprenolol 3.11/3.04 (2.89, 3.10, 3.13), betaxolol 2.81/2.81 (2.81), timolol 1.88/1.87 (1.83, 1.91) (Hansch and Leo, 1979; Hansch et al., 1995; Mannhold et al., 1990; Recanatini, 1992; Schoenwald and Huang, 1983). Higher pH values were avoided as the oxime function may not be very stable under strongly basic conditions and isomerization is also accelerated. In all cases, the oximes were found slightly more lipophilic than the corresponding parent compounds (alcohols). In addition, for all cases, including the methoximes, log *P*

of the Z isomer was found to be slightly higher (more lipophilic) than that of the E isomer, in agreement with the chromatographic retention data. For the only somewhat similar case found in the literature the situation is the same; log *P* for the syn (Z) isomer of benzaldoxime is somewhat higher than log *P* of the anti (E) isomer: 1.85 versus 1.79, respectively (Hansch et al., 1995).

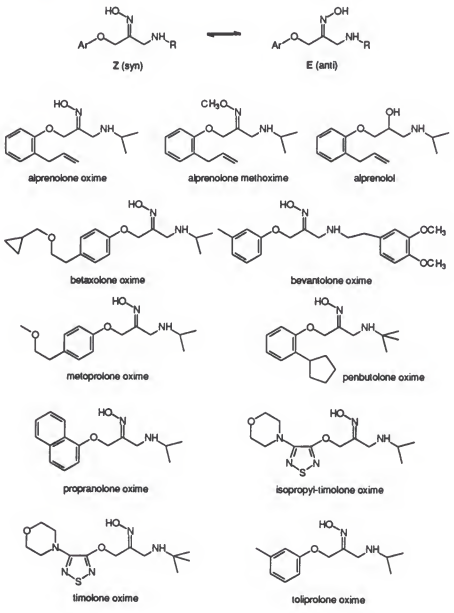


Figure 6.3 Structures for the compounds included in the present study. The two possible isomers (Z and E) are also represented for a general structure.

On the other hand, the log distribution coefficient (log *D*) measured at pH 7.0 was higher for all E isomers with 0.05-0.1 log units. Considering equation 6.2, this indicates a slightly, about 0.1-0.2 units, higher pK_a for the Z isomers. Therefore, they are somewhat more ionized at neutral pH. As this is true not only for the oximes but for the two included methoximes as well, it suggests that these Z isomers are somewhat more basic since the amine function is sterically less hindered by the oxime (methoxime) group. The pK_a values obtained with this estimation (8.3-8.7) are indeed about one unit smaller than those of the parent β-blocker alcohols. Table 6.1 summarizes the log *P* and log *D* values measured here together with the estimated pK_a values. The Z isomers proved to be more lipophilic but also more basic than the corresponding E isomers for all studied oximes and methoximes. Bevantolone oxime, the only molecule that bears a primary carbon and a larger substituent on its amine function, might be the only exemption. It not only has an inversion in log *P* and log *D* of the isomers (if HPLC elution order is not inverted), but also has a lower pK_a than the other oximes; a similar tendency was noted for the corresponding alcohol too (pK_a only 8.38 for bevantolol) (Schoenwald and Huang, 1983). The slightly different partition and distribution properties of the two possible isomers underline again the role of molecular conformation in determining physicochemical properties. It also worth noting that while log *P*s of the oximes are only slightly higher than those of the parent alcohols, log *D*s measured around physiological pH are more than one unit higher due to the pK_a differences. This might very well be one of

the factors causing the longer durations of action of these oximes and methoximes when compared to the parent compounds (Bodor, 1994; Prokai et al., 1995; Farag et al., 1997).

Table 6.1. Measured and Calculated Properties for Compounds Included in the Present Study.[a]

No.	Compound	log k'_{25}	log *D*	log *P*	log P_{calc}	acdLogP	BLOGP	pK_a est.
1	alprenolol	0.95	0.66	3.11	3.15	2.88	2.90	9.45
2	alprenolone oxime E	1.06	1.88	3.34	3.30	3.18	2.32	8.44
3	alprenolone oxime Z	1.19	1.84	3.40	3.30	3.18	2.22	8.55
4	betaxolol	0.97	0.41	2.81	3.26	2.66		9.40
5	betaxolone oxime E	1.07	1.60	3.20	3.41	2.95	2.39	8.59
6	betaxolone oxime Z	1.19	1.55	3.32	3.41	2.95	2.24	8.76
7	betaxolone methoxime E	1.45	2.37	3.63	3.87	3.53	2.36	8.24
8	betaxolone methoxime Z	1.51	2.16	3.83	3.87	3.53	2.16	8.66
9	bevantolone oxime E	1.15	2.42	3.32	3.91	3.60	1.40	7.84
10	bevantolone oxime Z	1.27	2.43	3.29	3.91	3.60	1.89	7.80
11	metoprolone oxime E	0.43	0.67	2.19	2.26	2.05	2.25	8.51
12	metoprolone oxime Z	0.55	0.59	2.30	2.26	2.05	2.12	8.70
13	penbutolone oxime E		3.38		4.53	4.46	2.18	
14	penbutolone oxime Z		3.23		4.53	4.46	1.95	
15	propranolone oxime E	1.09	1.98	3.57	3.29	3.39	2.21	8.58
16	propranolone oxime Z	1.13	1.91	3.61	3.29	3.39	2.17	8.69
17	*i*propyl-timolone oxime E	0.12			-0.03	1.20	1.19	
18	*i*propyl-timolone oxime Z	0.23	0.34	1.03	-0.03	1.20	1.36	7.59
19	timolol	0.21	-0.45	1.88	0.22	-0.15		9.33
20	timolone oxime E	0.31			0.41	1.55		
21	timolone oxime Z	0.39	0.44	1.99	0.41	1.55		8.54
22	timolone methoxime E	0.61	1.32	2.62	0.87	2.13		8.28
23	timolone methoxime Z	0.64	1.25	2.68	0.87	2.13		8.41
24	toliprolone oxime E	0.61	1.34	2.74	2.58	2.62	2.12	8.38
25	toliprolone oxime Z	0.76	1.29	2.76	2.58	2.62	2.00	8.46

[a] Included data are as follows: log k'_{25}: log chromatographic capacity factor obtained for a mobile phase containing 25% acetonitrile; log *D*: log distribution coefficient measured at pH 7.0; log *P*: log partition coefficient (for neutral compounds); log P_{calc}: calculated values using the molecular size-based approach; ACD/LogP: calculated values using the commercially available ACD/LogP package (Advanced Chemistry Development Inc., 141 Adelaide St. West, Suite1501, Toronto, Ontario, Canada); pK_a est.: estimated pK_a values using measured log *D* and log *P* data in equation 6.2.

Despite having acetonitrile as organic modifier in the mobile phase and not methanol, which is usually considered the most adequate cosolvent in establishing such type of relationships, capacity factors obtained for a mobile

phase containing 25% acetonitrile correlate quite well with obtained log *P* values, and they both give the same order of lipophilicity for different compounds and for E and Z isomers as well. Chromatographic capacity factors (k') were calculated as $k' = (t_r - t_0)/t_0$, where t_r represents the retention time and t_0 the dead time determined by injection of sodium nitrate solution.

$$\log P = 1.706(\pm 0.150) \log k'_{25} + 1.392(\pm 0.143);$$
$$n = 21,\ r = 0.933,\ \sigma = 0.258,\ F = 128.8 \qquad (6.3)$$

Calculated log *P*s also agree reasonably well with measured ones. The present molecular size-based approach mispredicts only timolol and related compounds. This is somehow unexpected, as other compounds having thiazole or morpholine constituents are not mispredicted, even if the thiazole moiety was not present in the database used to develop the model. ACD/LogP (Advanced Chemistry Development Inc., 141 Adelaide St. West, Suite1501, Toronto, Ontario, Canada); seriously mispredicts only timolol and not the derived oximes; however, being a fragment based method it does so by simply introducing new, special fragment values and corresponding interaction factors for these groups. Unfortunately, BLOGP, the only method able to distinguish between the two isomers, is unable to give a good correlation; for most cases it also mispredicts their relative lipophilicity, predicting a higher log *P* for the E isomer.

Partition data could not be determined for the E isomers of timolone and isopropyl timolone oximes, as they are rapidly converted to a different, yet unidentified compound. This conversion at pH 7.0 or higher is fast enough ($t_{1/2}$

~5 min.) not to allow for equilibration in partition measurements. No such problem occurred for the corresponding Z isomers, for the methoxime analogues, or for any other compound, except maybe for bevantolone oxime at a significantly slower rate.

Traditional shake-flask method with an RP-HPLC procedure developed to separate the E and Z isomers of the oxime or methoxime analogues of known β-adrenergic blockers was used to estimate the lipophilicity of these compounds. Experimental values agree well with calculated values, except timolol derivatives. Measured values indicate that the Z isomers are slightly more lipophilic than the corresponding E isomers. Additional distribution measurements at neutral pH indicate that the oximes are less basic than the corresponding parent aryloxy-alkylaminoalcohols and the E isomers, with a more hindered amine function, are somewhat less basic than the Z isomers. All these reemphasize the important role played by structural conformation and intramolecular interactions in determining physicochemical properties. The oximes have more than one log unit higher apparent partition coefficients at neutral pH than the parent β-adrenergic blocker alcohols; this might be one reason behind their longer duration of action.

Enzymatic Physical-Chemical–Based Chemical Delivery Systems

Design Principles

In this approach, the drug is chemically modified to introduce the protective function(s) and the targetor T moiety. Upon administration, the resulting CDS is distributed throughout the body. Predictable enzymatic reactions convert the original CDS by removing some of the protective functions and modifying the T moiety, leading to a precursor form, shown here as T^{+}-D, which is still inactive, but has significantly different physicochemical properties. While these intermediates are continuously eliminated from the "rest of the body," efflux-influx processes at the targeted site are not the same, due to the presence of a specific membrane or other distributional barrier, and they will provide a specific concentration here, ultimately allowing release of the active drug only at the site of action.

For example, the BBB can be regarded as a biological membrane that is permeable to most lipophilic compounds, but not to hydrophilic molecules, and in most cases, these transport criteria apply to both sides of the barrier. Thus, if a lipophilic CDS that can enter the brain is converted here to a hydrophilic molecule, one can assume that it will be "locked-in": it will no longer be able to come out. The same conversion taking place in the rest of the body actually accelerates peripheral elimination and further contributes to brain targeting.

In principle, many targetor moieties are possible for a general system of this kind, but the one developed some years ago (Bodor *et al.*, 1981; Bodor and

Brewster, 1983) based on the 1,4-dihydrotrigonelline ↔ trigonelline system proved the most useful. Here the T is this redox system where the lipophilic 1,4-dihydro form (T) is converted *in vivo* to the highly hydrophilic quaternary form (T^+). Since this system is closely related to the ubiquitous NADH ↔ NAD^+ coenzyme system, this conversion takes place very easily everywhere in the body. The resulting charged T^+ is locked into the brain, but is easily eliminated from the body due to the acquired positive charge. After a relatively short time, the drug D (as the inactive, locked-in T^+-D) is present essentially only in the brain, providing a sustained, brain-specific release of the active drug. The system is summarized in Figure 6.4, and it should be mentioned that selection of the nontoxic N-methylnicotinate (trigonelline) – dihydrotrigonelline system is also critical for its success. Examples of the use of this system for a wide variety of drug classes are abundant: various steroid hormones, anti-infective agents, antivirals, anticancer agents, antiretroviral agents (AZT, 2',3'-didehydro-2',3'-dideoxythymidine, etc.), neurotrophomodulator catechol derivatives and many others (Bodor, 1994; Bodor and Buchwald, 1997b; and references therein). Among these, estradiol-CDS is in the most advanced stage; it is already undergoing phase I and II clinical trials. In rats, estradiol (E_2) released from the T^+-E_2 intermediate formed after E_2-CDS administration has an elimination half-life of more than 200 hours, and brain E_2-levels are elevated four to five times longer after E_2-CDS administration than after simple estradiol treatment (Mullersman et al., 1988).

Most recently, successful brain deliveries of enkephalin (Bodor et al., 1992b; Prokai-Tatrai et al., 1996) and thyrotropin-releasing hormone (TRH) (Prokai et al., 1994) analogues using a combination of approaches, including the redox targetor system, were reported. Delivery of such peptides through the BBB is an even more complex problem, as BBB enzymes recognize and rapidly degrade most naturally occurring neuropeptides (Brownlees and Williams, 1993). Therefore, neither simple lipid soluble derivatives nor enzymatically stabilized variants, where natural L amino acids are replaced with D analogues, were successful. Thus, it became clear that three major issues must be solved simultaneously: enhance passive transport by increasing the lipophilicity to enable delivery, assure enzymatic stability to prevent premature degradation, and exploit the lock-in mechanism to provide targeting. This is accomplished with a complex *molecular packaging* strategy, a complex physical-chemical-enzymatic approach (Bodor and Prokai, 1995), where the peptide unit is part of a bulky molecule, dominated by lipophilic modifying groups that direct BBB penetration and prevent recognition by peptidases. In general, a brain-targeted packaged peptide delivery system contains the following major components: the redox T; a *spacer* function (S) consisting of strategically used amino acids to assure timely removal of the charged T (together with the S) from the peptide; the *peptide* itself (P); and a bulky *lipophilic moiety* (L) attached through an ester bond or sometimes through a C-terminal *adjuster* (A) at the carboxyl terminal to enhance lipid solubility and to disguise the peptide nature of the molecule.

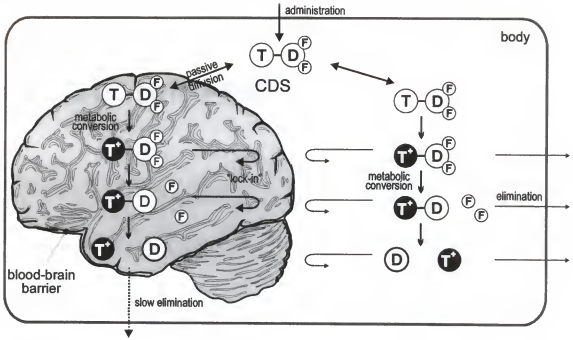


Figure 6.4 Schematic representation of the enzymatic physical-chemical-based CDS approach used for brain targeting. The original, inactive CDS is lipophilic, and is distributed throughout the body. Sequential metabolic conversions remove the modifier functions (F) and the targetor moiety (T), leading to a hydrophilic precursor (T^+-D) that is locked in the brain by the blood-brain barrier, but is easily eliminated from the rest of the body. The approach allows targeted and sustained delivery of the active drug (D).

To achieve delivery with such a complex system, it is vital that the designated enzymatic reactions take place in a specific sequence. Upon delivery, the first step must be the conversion of the T to allow for "lock-in." This is followed by removal of the L function to form a direct precursor of the peptide that is still attached to the charged targetor T^+. Subsequent cleavage of the T^+-S moiety finally leads to the active peptide.

The first successful delivery (Bodor et al., 1992), with a package named one of the top 10 medical advances of 1992 by the *Harvard Health Letter*, was

for Tyr-D-Ala-Gly-Phe-D-Leu (DADLE), an analogue of leucine-enkephalin, a naturally occurring linear pentapeptide (Tyr-Gly-Gly-Phe-Leu) that binds to opioid receptors. These opioid peptides are best-known for their central analgesic effect, but they are also involved in the regulation of seizure threshold, motor behavior, feeding, drinking, mental disorders, and immune responses, etc. In rat brain tissue samples collected 15 minutes after i.v. CDS administration, electrospray ionization mass spectrometry clearly showed the presence of the locked in T*-D form at an estimated concentration of 600 pmol per gram of tissue (Bodor et al., 1992). The same ion was absent from the sample collected from the control animals treated with the vehicle solution only. Electrospray ionization, the technique used in these mass spectrometric determinations, is capable of ionizing compounds with molecular weights up to 100 000 and puts many charges, in the form of protons, on certain large molecules such as polypeptides.

To optimize this delivery strategy, an effective synthetic route for peptide CDSs was established, and the role of the spacer and the lipophilic functions was investigated (Prokai-Tatrai et al., 1996). Intravenous injection of the four CDSs synthesized by a segment-coupling method produced a significant and long-lasting (more than 5 h) response in rats monitored with the tail-flick latency test. This measures the time until the animals, usually rats, withdraw their tails from a heated spot. When analgesics are administered, the animals are less likely to notice the pain caused by the heat, and the average withdrawal-time increases. The better the analgesic effect, the longerr the latency. The efficacy

of the CDS could be greatly influenced by modifying the spacer (S) and lipophilic (L) components, proving that they have important roles in determining molecular properties and timing of the metabolic sequence. The bulkier cholesteryl group used as L showed a better efficacy than the smaller 1-adamantaneethyl, but the most important factor to manipulate the rate of peptide release and the pharmacological activity turned out to be the S function: proline as spacer produced more potent analgesia than alanine.

A similar strategy was used to deliver a thyrotropin-releasing hormone (TRH) analogue to the central nervous system (Prokai et al., 1994). These analogues are potential agents for treating neurodegenerative disorders such as Alzheimer's disease (AD). However, because the "final" peptide following delivery has no free NH_2 and COOH termini, a precursor sequence (Gln-Leu-Pro-Gly) that will ultimately produce the "final" TRH analogue was packaged for delivery. This meant including two additional steps in the metabolic sequence: one where the C-terminal adjuster glycine is cleaved by peptidyl glycine α-amidating monooxygenase to form the ending prolinamide, and one where the N-terminal pyroglutamyl is formed from glutamine by glutaminyl cyclase. Selection of a suitable spacer was also important for the efficacy of TRH-CDSs as measured by the decrease in the barbiturate-induced sleeping time in mice after i.v. injection. At equimolar doses (30 μmol/kg), the TRH analogue itself showed only a marginal decrease, while the CDS with L-Ala spacer produced a

ca 30% reduction, and the one with a double alanine spacer produced a significant >50% decrease.

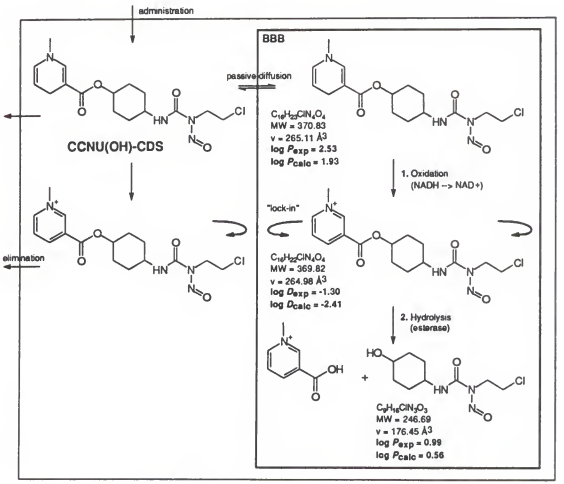


Figure 6.5 Schematic representation of the "lock-in" mechanism for the brain-targeting CDS of (hydroxy)lomustine [CCNU(OH)].The neutral, lipophilic CDS can easily cross the BBB, while the positively charged, hydrophilic intermediate is no longer able to come out, providing a sustained release of the active drug.

Physicochemical Data—Calculations and Measurements

Since, for these CDSs, adequate changes in physicochemical (i.e., partition) properties during the sequential metabolism are crucial for successful targeting, it is of considerable importance to acceptably predict such properties

starting from molecular structure. However, previously developed computer tools often cannot be used here. The latest CDSs are quite large molecules that are difficult to handle, even with AM1-based (Dewar et al., 1985) quantum chemical methods. This is especially true for peptide-CDSs, not only because they are large, but also because these methods have difficulties correctly describing peptide bonds. In addition, amino acids and peptides, an area of increasing pharmaceutical interest, still lack general, fast, and relatively reliable predictive methods. Most models developed for other molecules become quite unreliable here, while methods specifically developed for peptides cannot be used when nonpeptide constituents are present. Nevertheless, the distinctively simple model (Bodor and Buchwald, 1997), described in previous chapters, that estimates partition properties for a wide variety of organic molecules using a three-dimensional, molecular-size based approach proved useful here. By avoiding overparametrization, the model computes sound log *P* values even for some large compounds (e.g., cyclosporine, octreotide), where usual two-dimensional fragmental methods completely fail (Buchwald and Bodor, 1997, Bodor et al., 1997b). Distribution coefficients of charged species are very difficult to predict as they are counterion-dependent, but partition data of quaternary pyridinium-type compounds (Bodor and Brewster, 1983; Brewster et al., 1988; Hansch et al., 1995) around physiological pH could be reasonably well described with the present model by using a value of $\pi = 9$ for quaternary pyridinium nitrogens: $n = 21$, $r = 0.940$, $\sigma = 0.514$, $F = 143.74$ (Table 6.2).

Compared to the neutral form (π = 2 for pyridine), this represents a significant (7×0.72 ≈ 5) change, and such a change of around five orders of magnitude in partition properties could easily explain why the "lock in" mechanism can be successful. The mechanism, together with structures, metabolic conversions, calculated and measured physicochemical properties are summarized for brain-targeting (hydroxy)lomustine-CDS and estradiol-CDS in figures 6.5 and 6.6, respectively.

Table 6.2 Partition Data For Compounds Containing Quaternary Pyridinium-Type Moieties.

Compound	log *D*	log D_{calc}
N-Methylpyridinium (iodide)	-3.30	-3.94
3-Bromo methylpyridinium (iodide)	-2.91	-3.14
Butylpyridinium (bromide)	-2.69	-2.58
Hexylpyridinium (bromide)	-2.03	-1.69
Octylpyridinium (bromide)	-0.95	-0.80
β-Phenylethylpyridinium (bromide)	-2.35	-1.57
Decylpyridinium (bromide)	-0.72	0.09
N-Methyl-4-(4'-dimethylaminostyryl)-pyridinium (chloride)	-0.92	-0.82
N-Methyl-4-(4'-diethylaminostyryl)-pyridinium (chloride)	-0.66	0.08
N-Methyl-4-(4'-dipropylaminostyryl)-pyridinium (chloride)	-0.26	0.97
N-Methyl-4-(4'-dibutylaminostyryl)-pyridinium (chloride)	0.62	1.86
N-Methyl-4-(4'-dipentylaminostyryl)-pyridinium (chloride)	1.88	2.76
Pralidoxime [2((Hydroxyimino)Me)-1-methylpyridinium] (sulfonate)	-3.24	-3.82
PAD [2((Hydroxyimino)Me)-1-dodecylpyridinium] (iodide)	1.41	1.08
Berberine	-1.66[a]	-1.66
(Hydroxy)lomustine-CDS$^+$ [CCNU(OH)-CDS$^+$] (iodide)	-1.30	-2.41
AZT-CDS$^+$ (iodide)	-2.00	-3.65
Ethisterone-CDS$^+$ (iodide)	-0.03	0.33
Norethindrone-CDS$^+$ (iodide)	0.02	-0.08
Estradiol-CDS$^+$ (iodide)	-0.14	-0.09
TRH-CDS$^+$ (S = Pro-Pro, L = Cholesteryl)	0.25	0.82

[a]average of two values from Bodor and Brewster (1983), respectively Hansch et al. (1995).

Prediction for large and special compounds always have to be treated with caution, but we have one example of brain-targeting peptide CDS where experimental measurements confirmed our prediction. The distribution

coefficient at pH 7.0 (potassium phosphate monobasic sodium hydroxide buffer, 0.01 M) could be determined with traditional shake-flask method for the quaternary form of a CDS used to successfully deliver a thyrotropin-releasing hormone (TRH) analogue to the brain (Figure 6.7). Concentration measurements in both phases using both UV spectrometry and HPLC procedures indicated log $D = 0.25$ in good agreement with the predicted 0.82. Structures, physicochemical properties, and metabolic conversions are summarized in Figure 6.7 for such a system designed to deliver a TRH analogue to the CNS.

In summary, a computer program that uses a 3D molecular size-based approach of remarkable simplicity to estimate partition properties for organic compounds including neutral peptides was developed. As any reasonable (e.g., molecular mechanics optimized) three-dimensional structure can be used as a starting point and all algorithms are sufficiently fast, predictions can be obtained in a relatively fast and easy manner.

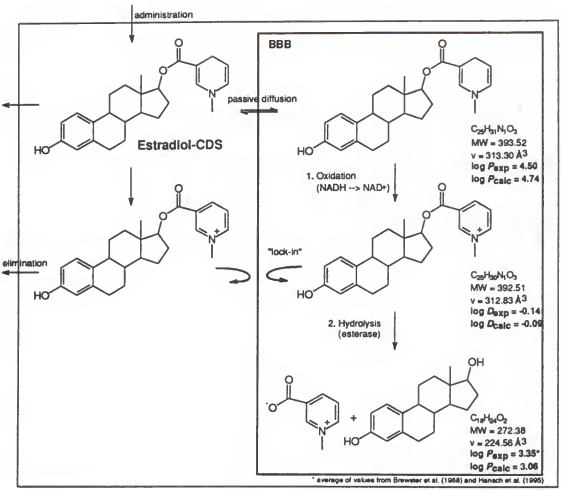


Figure 6.6 Illustration of the "lock-in" mechanism for estradiol-CDS. Log *P* and distribution coefficient (log *D*) are shown to illustrate the significant changes in partition properties. The lipophilic CDS (log $P > 4$) can easily cross the BBB, while the hydrophilic intermediate (log $D < 0$) is no longer able to come out, providing a sustained release of the active estradiol.

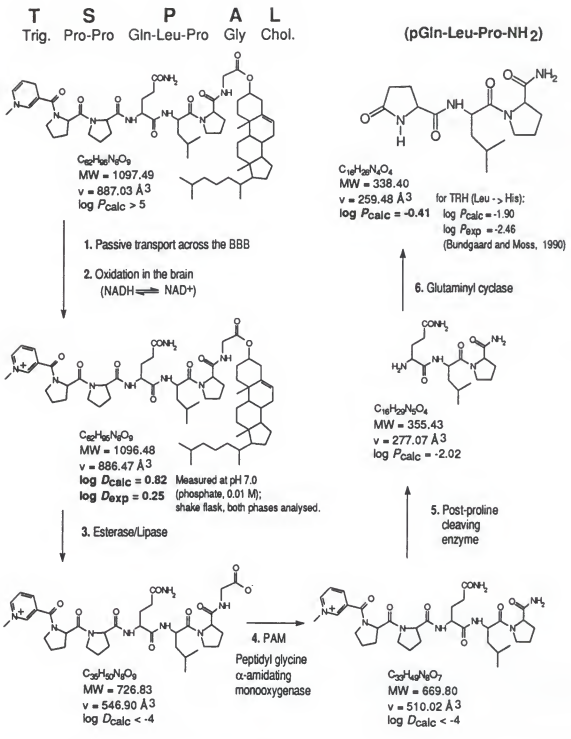


Figure 6.7 Sequential metabolism for a brain-targeting CDS of a TRH analogue and data describing the change in size and partition properties. The molecular package designed for brain-targeted peptide delivery contains a redox T, an S function, the peptide itself (P), a C-terminal adjuster (A), and a bulky L. Calculated values shown were obtained with the fully computerized method, as described in the text.

CHAPTER 7
CONCLUSIONS

The presented chapters together could represent the basis for a consistent approach relying on a very limited number of parameters to describe and to estimate physicochemical properties for organic liquids, an area of special pharmaceutical interest. Using a molecular size-based free energy approach, unified and consistent theoretical treatment was possible. Large number of applications and some experimental determinations were also included.

The first algorithm gives sufficiently reliable intrinsic size estimates much faster than entirely numerical programs. Volumes and surfaces obtained this way can be useful in different QSAR, molecular association, receptor binding, and other studies.

From a medicinal chemist's or drug designer's point of view, the most important part is the log *P* predictive model with a corresponding, fully implemented software. It successfully predicts log octanol-water partition coefficients (log *P*), probably the most informative and successful physicochemical property in medicinal chemistry, for a very wide variety of organic molecules using only two parameters. The computer program is under continuous development, and it can become a useful tool in drug design. A remarkable economy in parameters clearly distinguishes this method from usual

fragment based models, yet it works well enough for a wide range of organic compounds, including many pharmaceutically important molecules. It also gives some insight into the behavior of "two-dimensional" fragment approaches.

As a noteworthy achievement, extension of the same, molecular size-based model to predict log *P* of neutral peptides was possible. Despite the use of only two parameters, promising correlations were obtained, and predictions were in a realistic range even for larger peptides (cyclosporin, octreotide, gramicidin) where common, overparametrized fragment methods completely fail. After a brief review of the theoretical background of ion pair partitioning, a possible extension of the model was also performed to give a first estimate of distribution coefficients measured around physiological conditions. Physicochemical considerations that were obtained from the present model and some experimental determinations and relate to the retrometabolic drug design of brain- and eye-targeting chemical delivery systems are included in a separate chapter. A standard experimental method for log $P_{o/w}$ determinations was developed and tested; some new data are reported.

The model for organic liquid properties is in a preliminary stage; some assumptions still need additional validation. The described version works on a large number but only relatively simple molecules. Nevertheless, it was included because it made possible a consistent and systematic approach. Such an approach makes unnecessary to rely on often misused concepts of standard states and activities, and it leads to relations that can not be obtained from

purely thermodynamic arguments. Starting from molecular assumptions and using only van der Waals molecular volume as predictor, it gives a unified description of boiling points, enthalpies of vaporizations, and partition coefficients for liquids having no strongly polar or hydrogen bonding substituents accounting for 90-95% of the variance in these properties. The possibility to use an interaction constant obtained from enthalpies of vaporizations and boiling points to describe gas-liquid and possibly liquid-liquid partitions might become the most important achievement of the present work.

REFERENCES

Abraham, D. J.; Leo, A. J. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins* **1987**, *2*,130-152.

Abraham, M. H. Free energies of solution of rare gases and alkanes in water and nonaqueous solvents. A quantitative assessment of the hydrophobic effect. *J. Am. Chem. Soc.* **1979**, *101*, 5477-5484.

Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* **1994b**, *83*, 1257-1268.

Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Δlog P parameter of Seiler. *J. Pharm. Sci.* **1994**a, *83*, 1085-1100.

Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. Thermodynamics of solute transfer from water to hexadecane. *J. Chem. Soc. Perkin Trans. 2* **1990**, 291-300.

Akamatsu, M.; Fujita, T. Quantitative analyses of hydrophobicity of di- to pentapeptides having un-ionizable side chains with substituent and structural parameters. *J. Pharm. Sci.* **1992**, *81*, 164-173,.

Akamatsu, M.; Katayama, T.; Kishimoto, D.; Kurokawa, Y.; Shibata, H.; Ueno, T.; Fujita, T. Quantitative analyses of the structure-hydrophobicity relationship for N-acetyl di- and tripeptide amides. *J. Pharm. Sci.* **1994**, *83*,1026-1033.

Alberty, R. A. *Physical Chemistry*, 7th edn., John Wiley & Sons: New York, 1987.

Alberty, R. A.; Silbey, R. J. *Physical Chemistry*, 2nd edn., Wiley: New York, 1997; p 950.

Alejandre, J.; Martinez-Casas, S. E.; Chapela, G. A. On the calculation of the sphericity factor for fused hard sphere molecules. *Mol. Phys.* **1988**, *65*, 1185-1193.

Alkorta, I.; Villar, H. O. Quantum mechanical parametrization for a conformationally dependent hydrophobic index. *Int. J. Quantum Chem.* **1992**, *44*, 203-218.

Amidon, G. L.; Yalkowsky, S. H.; Anik, S. T.; Valvani, S. C. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach. *J. Phys. Chem.* **1975**, *79*, 2239-2246.

Audry, E.; Dubost, J.-P.; Colleter, J.-C.; Dallet, P. Une nouvelle approche des relations structure-activité: le "potentiel de lipophilie moléculaire". *Eur. J. Med. Chem.* **1986**, *21*, 71-72.

Austin, R. P.; Davis, A. M.; Manners, C. N. Partitioning of ionizing molecules between aqueous buffers and phospholipid vesicles. *J. Pharm. Sci.* **1995**, *84*, 1180-1183.

Avdeef, A.; Box, K. J.; Takács-Novák, K. pH-Metric logP. 6. Effects of sodium, potassium, and N-CH_3-D-glucamine on the octanol-water partitioning of prostaglandins E_1 and E_2. *J. Pharm. Sci.* **1995**, *84*, 523-529.

Bader, R. F. W.; Carrol, M. T.; Cheeseman, J. R.; Chang, C. Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.* **1987**, *109*, 7968-7979.

Banks, W. A.; Kastin, A. J. Peptides and the blood-brain barrier: Lipophilicity as a predictor of permeability. *Brain Res. Bull.* **1985**, *15*, 287-292.

Begley, D. J. Peptides and the blood-brain barrier. In *Physiology and Pharmacology of the Blood-Brain Barrier, Handbook of Experimental Pharmacology*, vol. 103, Bradbury, M. W. B., Ed.; Springer Verlag: Berlin, 1992; pp 151-203.

Begley, D. J. The blood-brain barrier: Principles for targeting peptides and drugs to the central nervous system. *J. Pharm. Pharmacol.* **1996**, *48*, 136-146.

Biagi, G. L.; Barbaro, A. M.; Guerra, M. C.; Babbini, M.; Gaiardi, M.; Bartoletti, M.; Borea, P. A. R_m values and structure-activity relationship of benzodiazepines. *J. Med. Chem.* **1980**, *23*, 193-201.

Ben-Naim, A. *Hydrophobic Interactions*, Plenum Press: New York, 1980; p 260.

Ben-Naim, A. Standard thermodynamics of transfer: Uses and misuses. *J. Phys. Chem.* **1978**, *82*, 792-803.

Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016-2027.

Bliznyuk, A. A.; Gready, J. E. Numerical calculation of molecular surface area. I. Assessment of errors. *J. Comput. Chem.* **1996**, *17*, 962-969.

Bock, M. G.; DiPardo, R. M.; Evans, B. E.; Freidinger, R. M.; Rittle, K. E.; Payne, L. S.; Boger, J.; Whitter, W. L.; LaMont, B. I.; Ulm, E. H.; Blaine, E. H.; Schorn, T. W.; Veber, D. F. Renin inhibitors containing hydrophilic groups. Tetrapeptides with enhanced aqueous solubility and nanomolar potency. *J. Med. Chem.* **1988**, *31*, 1918-1923.

Bodor, N. Redox drug delivery systems for targeting drugs to the brain. *Ann. NY Acad. Sci.* **1987**, *507*, 289-306.

Bodor, N. Drug targeting and retrometabolic drug design approaches. *Adv. Drug Delivery Rev.* **1994**, *14*, 157-166.

Bodor, N. Retrometabolic approaches to drug targeting. In *NIDA Research Monograph Series*, vol. 154., *Membranes and Barriers: Targeted Drug Delivery*. Rapaka, R., Ed.; NIDA: Rockville, MD, 1995a; pp 1-27.

Bodor, N. Retrometabolic drug design concepts in ophthalmic target-specific drug delivery. *Adv. Drug Delivery Rev.* **1995b**, *16*, 21-38.

Bodor, N.; Brewster, M. E. Problems of delivery of drugs to the brain. *Pharmacol. Ther.* **1983**, *19*, 337-386.

Bodor, N.; Buchwald, P. Drug targeting via retrometabolic approaches. *Pharmacol. Ther.* **1997b**, *76*, in press.

Bodor, N.; Buchwald, P. Molecular size based approach to estimate partition properties for organic solutes. *J. Phys. Chem. B* **1997a**, *101*, 3404-3412.

Bodor, N.; Buchwald, P.; Huang, M.-J. Computer assisted design of new drugs based on retrometabolic concepts. *SAR QSAR Environ. Res.* **1997b**, in press.

Bodor, N.; El-Koussi, A.; Kano, M.; Nakamuro, T. Improved delivery through biological membranes. 26. Design, synthesis, and pharmacological activity of a novel chemical delivery system for β-adrenergic blocking agents. *J. Med. Chem.* **1988**, *31*, 100-106.

Bodor, N.; Farag, H. H.; Brewster, M. E. Site-specific, sustained release of drugs to the brain. *Science* **1981**, *214*, 1370-1372.

Bodor, N.; Farag, H. H.; Somogyi, G.; Wu, W.-M.; Barros, M. D. C.; Prokai, L. Ocular-specific delivery of timolol by sequential bioactivation of its oxime and methoxime analogs. *J. Ocul. Pharmacol.* **1997a**, *13*, 389-403.

Bodor, N.; Gabanyi, Z.; Wong, C.-K. A new method for the estimation of partition coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783-3786.

Bodor, N.; Huang, M.-J. Intermolecular interactions of methyl acetate, β-propiolactone, ethyl acetate and γ-butyrolactone. An AM1 semi-empirical study. *Int. J. Quant. Chem.* **1992a**, *44*, 81-89.

Bodor, N.; Huang, M.-J. An extended version of a novel method for the estimation of partition coefficients. *J. Pharm. Sci.* **1992b**, *81*, 272-281.

Bodor, N.; Huang, M.-J. A new method for the estimation of the aqueous solubility of organic compounds. *J. Pharm. Sci.* **1992c**, *81*, 954-960.

Bodor, N.; Huang, M.-J.; Harget, A. Neural network studies. Part 3. Prediction of partition coefficients. *J. Mol. Struct. (Theochem)* **1994**, *309*, 259-266.

Bodor, N.; Kaminski, J. J. Prodrugs and site-specific chemical delivery systems. *Annu. Rep. Med. Chem.* **1987**, *22*, 303-313.

Bodor, N.; Prokai, L. Molecular packaging: peptide delivery to the central nervous system by sequential metabolism. In *Peptide-Based Drug Design: Controlling Transport and Metabolism.* Taylor, M., Amidon, G., Eds.; American Chemical Society: Washington, DC, 1995; pp 317-337.

Bodor, N.; Prokai, L. Site- and stereospecific ocular drug delivery by sequential enzymatic bioactivation. *Pharm. Res.* **1990**, *7*, 723-725.

Bodor, N.; Prokai, L.; Wu, W-M.; Farag, H. H.; Jonnalagadda, S.; Kawamura, M.; Simpkins, J. A strategy for delivering peptides into the central nervous system by sequential metabolism. *Science* **1992**, *257*, 1698-1700.

Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441-451.

Brewster, M. E.; Estes, K. S.; Bodor, N. Improved delivery through biological membranes. XXXII. Synthesis and biological activity of brain-targeted delivery systems for various estradiol derivatives. *J. Med. Chem.* **1988**, *31*, 244-249.

Broto, P.; Moreau, G.; Vandycke, C. Molecular structures: perception, autocorrelation descriptor and SAR studies. System of atomic contributions for the calculation of the n-octanol/water partition coefficients. *Eur. J. Med. Chem.* **1984**, *19*, 71-78.

Brownlees, J.; Williams, C. H. Peptidases, peptides, and the mammalian blood-brain barrier. *J. Neurochem.* **1993**, *60*, 793-803.

Buchwald, P.; Bodor, N. Octanol-water partition of nonzwitterionic peptides: Predictive power of a molecular size-based model. *Proteins* **1997**, in press.

Bundgaard, H., Design of prodrugs: Bioreversible derivatives for various functional groups and chemical entities. In *Design of Prodrugs*, Bundgaard, H., Ed.; Elsevier: Amsterdam, 1985; pp 1-92.

Bundgaard, H.; Moss, J. Prodrugs of peptides. 6. Bioreversible derivatives of thyrotropin-releasing hormone (TRH) with increased lipophilicity and resistance to cleavage by the TRH-specific serum enzyme. *Pharm. Res.* **1990**, *7*, 885-892.

Burton, P. S.; Conradi, R. A.; Hilgers, A. R. Transcellular mechanism of peptide and protein absorption: passive aspects. *Adv. Drug Del. Rev.* **1991**, *7*, 365-386.

Burton, P. S.; Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Maggiora, L. L. The relationship between peptide structure and transport across epithelial cell monolayers. *J. Controlled Release* **1992**, *19*, 87-98.

Calvino, R.; Gasco, A.; Leo, A. An analysis of the lipophilicity of furazan and furoxan derivatives using CLOGP algorithm. *J. Chem. Soc. Perkin Trans. II* **1992**, 1643-1646.

Cefalu, W. T.; Pardridge, W. M. Restrictive transport of a lipid-soluble peptide (cyclosporin) through the blood-brain barrier. *J. Neuroch.* **1985**, *45*, 1954-1956.

Chikhale, E.G.; Ng, K.-Y.; Burton, P.S.; Borchardt, R.T. Hydrogen bonding potential as a determinant of the in vitro and in situ blood-brain barrier permeability of peptides. *Pharm. Res.* **1994**, *11*, 412-419.

Chou, J. T.; Jurs, P. C. Computer-assisted computation of partition coefficient from molecular structure using fragment constants. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 172-178.

Collander, R. The partition of organic compounds between high alcohols and water. *Acta Chem. Scand.* **1951**, *5*, 774-780.

Collander, R. The permeability of plant protoplasts to nonelectrolytes. *Trans. Faraday Soc.* **1937**, *33*, 985-990.

Collander, R.; Bärlund, H. Permeabilitätsstudien an *Chara ceratophylla*. II. Die Permeabilität für Nichtelektrolyte. *Acta Botan. Fennica* **1933**, *11*, 1-114.

Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Cryst.* **1983**, *16*, 548-558.

Connolly, M. L. Computation of molecular volume. *J. Am. Chem. Soc.* **1985**, *107*, 1118-1124.

Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. II. Peptide bond modification which results in improved permeability. *Pharm. Res.* **1992**, *9*, 435-439.

Convard, T.; Dubost, J. P.; Le Solleu, H.; Kummer, E. Smilogp - a program for a fast evaluation of theoretical log-P from the Smiles code of a molecule. *Quant. Struct. Act. Relat.* **1994**, *13*, 34-37.

Cotton, R.; Hardy, P. M.; Langran-Goldsmith, A. E. Peptides containing dialkylglycines. *Int. J. Peptide Protein Res.* **1986**, *28*, 230-244.

Cramer, R. D., III. "Hydrophobic interaction" and solvation energies: Discrepancies between theory and experimental data. *J. Am. Chem. Soc.*, **1977**, *99*, 5408-5412.

Cramer, R. D., III. BC(DEF) Parameters. 1. The intrinsic dimensionality of intermolecular interactions in the liquid state. *J. Am. Chem. Soc.* **1980**, *102*, 1837-1849.

CRC Handbook of Chemistry and Physics, 72nd edn.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1991-1992.

Creamer, T. P.; Rose, G. D. α-Helix-forming propensities in peptides and proteins. *Proteins* **1994**, *19*, 85-97.

Debnath, A. K.; de Compadre, R. L. L.; Debnath, G.; Shusterman, A. J.; Hansch, C. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity. *J. Med. Chem.* **1991**, *34*, 786-797.

Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.

DeYoung, L. R.; Dill, K. A. Partitioning of nonpolar solutes into bilayers and amorphous n-alkanes. *J. Phys. Chem.* **1990**, *94*, 801-809.

do Carmo, M. P. *Differential Geometry of Curves and Surfaces*; Prentice-Hall: Englewood Cliffs, NJ, 1976; p 503.

Dodd, L. R.; Theodorou, D. N. Analytical treatment of the volume and surface area of molecules formed by an arbitrary collection of unequal spheres intersected by planes. *Mol. Phys.* **1991**, *72*, 1313-1345.

Dreisbach, D. *Liquids and Solutions*, Houghton Mifflin: Boston, 1966; p 194.

Drewe, J.; Fricker, G.; Vonderscher, J.; Beglinger, C. Enteral absorption of octreotide: Absorption enhancement by polyoxyethylene-24-cholesterol ether. *Br. J. Pharmacol.* **1993**, *108*, 298-303.

Dunn, W. J., III.; Grigoras, S.; Johansson, E. Principal component analysis of partition coefficient data: An approach to a new method of partiton coefficient estimation. In *Partition Coefficient Determination and Estimation*, Dunn, W. J., III., Block, J. H., Pearlman, R. S., Eds.; Pergamonn: New York, 1986; pp 21-36.

Dunn, W. J., III.; Nagy, P. I. Relative logP and solution structure for small organic solutes in the chloroform/water system using Monte Carlo methods. *J. Comput. Chem.* **1992**, *13*, 468-477.

Dunn, W. J., III.; Nagy, P. I.; Collantes, E. R. A computer assisted method for estimation of the partition coefficient. Monte Carlo simulations of the chloroform/water logP for methylamine, methanol and acetonitrile. *J. Am. Chem. Soc.* **1991**, *113*, 7898-7902.

Eisenhaber, F.; Argos, P. Improved strategy in analytical surface calculation for molecule systems: Handling singularities and computational efficiency. *J. Comput. Chem.* **1993**, *14*, 1272-1280.

El Tayar, N.; Mark, A. E.; Vallat, P.; Brunne, R. M.; Testa, B.; van Gusteren, W. F. Solvent dependent conformation and hydrogen-bonding capacity of cyclosporin A: Evidence from partition coefficients and molecular dynamics simulations. *J. Med. Chem.* **1993**, *36*, 3757-3764.

El Tayar, N.; Testa, B.; Carrupt, P.-A. Polar intermolecular interactions encoded in partition coefficients: an indirect estimation of hydrogen-bond parameters of polyfunctional solutes. *J. Phys. Chem.* **1992**, *96*, 1455-1459.

Famini, G. R.; Penski, C. A.; Wilson, L. Y. Theoretical descriptors in quantitative structure activity relationships: some physicochemical properties. *J. Phys. Org. Chem.* **1992**, *5*, 395-408.

Farag, H. H.; Wu, W.-M.; Barros, M. D. C.; Somogyi, G.; Prokai, L.; Bodor, N. Ocular-specific chemical delivery system of betaxolol for safe local treatment of glaucoma. *Drug Des. Discov.* **1997**, in press.

Fauchère, J.-L.; Pliska, V. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **1983**, *18*, 369-375.

Fauchère, J.-L.; Quarendon, P.; Kaetter, L. Estimating and representing hydrophobicity potential. *J. Mol. Graphics* **1988**, *6*, 203-206.

Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **1985**, *314*, 235-238.

Flynn, G. L.; Yalkowsky, S. N. Correlation and prediction of mass transport across membranes. I: Influence of alkyl chain length on flux-determining properties of barrier and diffusant. *J. Pharm. Sci.* **1972**, *61*, 838-852.

Franks, N. P.; Lieb, W. R. Where do general anaesthetics act? *Nature* **1978**, *274*, 339-342.

Fujita, T.; Iwasa, J.; Hansch, C. A new substituent constant, π, derived from partition coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175-5180.

Furet, P.; Sele, A.; Cohen, N. C. 3D molecular lipophilicity potential profiles: a new tool in molecular modeling. *J. Mol. Graphics* **1988**, *6*, 182-189.

Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular lipophilicity potential, a tool in 3D-QSAR. Method and applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83-96.

Gavezzotti, A. The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. *J. Am. Chem. Soc.* **1983**, *105*, 5220-5225.

Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. I. Partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565-577.

Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. III. Modeling hydrophobic interactions. *J. Comput. Chem.* **1988**, *9*, 80-90.

Gibson, K. D.; Scheraga, H. A. Exact calculation of the volume and surface area of fused hard-sphere molecules with unequal atomic radii. *Mol. Phys.* **1987**, *62*, 1247-1265.

Gibson, K. D.; Scheraga, H. A. Surface area of the intersection of three spheres with unequal radii. A simplified analytical formula. *Mol. Phys.* **1988**, *64*, 641-644.

Grant, D. J. W.; Higuchi, T. *Solubility Behavior of Organic Compounds*; Techniques of Chemistry, vol. 21, Wiley-Interscience: New York, 1990; p 600.

Greig, N. H. Drug entry into the brain and its pharmacologic manipulation. In *Physiology and Pharmacology of the Blood-Brain Barrier, Handbook of Experimental Pharmacology*, vol. 103, Bradbury, M. W. B., Ed.; Springer Verlag: Berlin, 1992; pp 487-523.

Guggenheim, E. A. On the statistical mechanics of dilute and perfect solutions. *Proc. Roy. Soc. London, Series A* **1932**, *135*, 181-192.

Haeberlein, M.; Brinck, T. Prediction of water-octanol partition coefficients using theoretical descriptors derived from the molecular surface area and the electrostatic potential. *J. Chem. Soc., Perkin Trans. 2* **1997**, 289-294.

Handbook of Data on Common Organic Compounds, vol. 1-3, eds. Lide, D. R.; Milne, G. W. CRC Press: Boca Raton, Florida,1995.

Hansch, C.; Björkroth, J. P.; Leo, A. Hydrophobicity and central nervous system agents: on the principle of minimal hydrophobicity in drug design. *J. Pharm. Sci.* **1987**, *76*, 663-687.

Hansch, C.; Clayton, J. M. Lipophilic character and biological activity of drugs II. The parabolic case. *J. Pharm. Sci.* **1973**, *62*, 1-21.

Hansch, C.; Dunn, J., III. Linear relationships between lipophilic character and biological activity of drugs. *J. Pharm. Sci.* **1972**, *61*, 1-19.

Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley-Interscience: New York, 1979; p 339.

Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR*. vol. 2., *Hydrophobic, Electronic, and Steric Constants*. American Chemical Society: Washington, DC, 1995; p 557.

Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **1962**, *194*, 178-180.

Hansch, C.; Steward, A. R.; Anderson, S. M.; Bentley, D. The parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. *J. Med. Chem.* **1968**, *11*, 1-21.

Harris, M. J.; Higuchi, T.; Rytting, J. H. Thermodynamic group contributions from ion pair extraction equilibria for use in the prediction of partition coefficients.

Correlation of surface area with group contributions. *J. Phys. Chem.* **1973**, *77*, 2694-2703.

Heiden, W.; Moeckel, G.; Brickmann, J. A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J. Comp. Aided Design* **1993**, *7*, 503-514.

Hermann, R. B. Modeling hydrophobic solvation of nonsperical systems: comparison of use of molecular surface area with accesible surface area. *J. Comp. Chem.* **1997**, *18*, 115-125.

Hermann, R. B. Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* **1972**, *76*, 2754-2759.

Higo, J.; Go, N. Algorithm for rapid calculation of excluded volume of large molecules. *J. Comput. Chem.* **1989**, *10*, 376-379.

Hildebrand, J. H.; Scott, R. L. *The Solubility of Nonelectrolytes*, Reinhold Publishing Corporation: New York, 1950.

Hopfinger, A. J.; Battershell, R. D. Application of SCAP to Drug Design. 1. Prediction of octanol-water partition coefficients using solvent-dependent conformational analyses. *J. Med. Chem.* **1976**, *19*, 569-573.

Hunter, E. B.; Powers, S. P.; Kost, L. J.; Pinon, D. I.; Miller, L. J.; LaRusso, N. F. Physicochemical determinants in hepatic extraction of small peptides. *Hepatology* **1990**, *12*, 76-82.

Iijima, H.; Dunbar, J. B., Jr.; Marshall, G. R. Calibration of effective van der waals contact radii for proteins and peptides. *Proteins* **1987**, *2*, 330-339.

Itokawa, H.; Takeya, K.; Mori, N.; Sonobe, T.; Serisawa, N.; Hamanaka, T.; Mihashi, S. Studies on antitumor cyclic hexapeptides RA obtained from Rubiae Radix, Rubiaceae. III. On derivatives of RA-V and their in vivo activities. *Chem. Pharm. Bull.* **1984**, *32*, 3216-3226.

Iwase, K.; Komatsu, K.; Hirono, S.; Nakagawa, S.; Moriguchi, I. Estimation of hydrophobicity based on the solvent-accesible surface area of molecules. *Chem. Pharm. Bull.* **1985**, *33*, 2114-2121.

Jorgensen, W. L.; Briggs, J. M.; Contreras, M. L. Relative partition coefficients for organic solutes from fluid simulation. *J. Phys. Chem.* **1990**, *94*, 1683-1686.

Kahns, A. H.; Buur, A.; Bundgaard, H. Prodrugs of peptides. 18. Synthesis and evaluation of various esters of desmopressin (dDAVP). *Pharm. Res.* **1993**, *10*, 68-74.

Kamlet, M. J.; Abboud, J.-L. M.; Abraham, M. H.; Taft, R. W. Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π^*, α, and β, and some methods for simplifying the generalized solvatochromic equation. *J. Org. Chem.* **1983**, *48*, 2877-2887.

Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Carr, P. W.; Doherty, R. F.; Taft, R. W. Linear solvation energy relationships. 41. Important differences between aqueous solubility relationships for aliphatic and aromatic solutes. *J. Phys. Chem.* **1987**, *91*, 1996-2004.

Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Marcus, Y.; Taft, R. W. Linear solvation energy relationships. 46. An improved equation for correlation and prediction of octanol/water partiton coefficients of organic nonelectrolytes (including strong hydrogen bond donor solutes). *J. Phys. Chem.* **1988**, *92*, 5244-5255.

Kantola, A.; Villar, H. O.; Loew, G.H. Atom based parametrization for a conformationally dependent hydrophobic index. *J. Comput. Chem.* **1991**, *12*, 681-689.

Karickhoff, S. W.; Brown, D. S.; Scott, T. A. Sorption of hydrophobic pollutants on natural sediments. *Water Res.* **1979**, *13*, 241-248.

Katayama, T.; Nakao, K.; Akamatsu, M.; Ueno, T.; Fujita, T. Quantitative structure-hydrophobicity and structure-activity relationships of antibacterial gramicidin S analogs. *J. Pharm. Sci.* **1994**, *83*, 1357-1362.

Kellogg, G. E.; Semus S. F.; Abraham, D. J. A new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aided Mol. Design* **1991**, *5*, 545-552.

Kihlberg, J.; Åhman, J.; Walse, B.; Drakenberg, T.; Nilsson, A.; Söderberg-Ahlm, C.; Bengtsson, B.; Olsson, H. Glycosylated peptide hormones: Pharmacological properties and conformational studies of analogues of [1-desamino, 8-D-arginine]vasopresssin. *J. Med. Chem.* **1995**, *38*, 161-169.

Kim, A.; Szoka, F. C., Jr. Amino acid side-chain contributions to free energy of transfer from water to octanol. *Pharm. Res.* **1992**, *9*, 504-514.

Klopman, G.; Iroff, L. Calculation of partition coefficients by the charge density method. *J. Comput. Chem.* **1981**, *2*, 157-160.

Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer automated log *P* calculations based on an extended group contribution approach. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752-781.

Klopman, G.; Wang, S. A computer automated structure evaluation (CASE) approach to calculation of partition coefficient. *J. Comput. Chem.* **1991**, *12*, 1025-1032.

Kossoy, A. D.; Risley, D. S.; Kleyle, R. M.; Nurok, D. Novel computational method for the determination of partition coefficients by planar chromatography. *Anal. Chem.* **1992**, *64*, 1345-1349.

Krauss, E.M.; Chan, S.I. Intramolecular hydrogen bonding in gramicidin S. 2. Ornithine. *J. Am. Chem. Soc.* **1982**, *104*, 6953-6961.

Laidler, K. J.; Meiser, J. H. *Physical Chemistry*, 2nd edn., Houghton Mifflin: Boston, 1995; p 988.

Lan, E.-L.; Ugwu, S. O.; Blanchard, J.; Fang, X.; Hruby, V. J.; Sharma, S. Preformulation studies with melanotan II: A potential skin cancer chemopreventive peptide. *J. Pharm. Sci.* **1994**, *83*,1081-1084.

Langs, D. A.; Smith, G. D.; Stezowski, J. J.; Hughes, R. E. Structure of pressinoic acid: The cyclic moiety of vasopressin. *Science* **1986**, *232*, 1240-1242.

Le Grand, L. M.; Merz, K. M., Jr. Rapid approximation to molecular surface area via the use of boolean logic and look-up tables. *J. Comput. Chem.* **1993**, *14*, 349-352.

Leahy, D. E. Intrinsic molecular volume as a measure of the cavity term in linear solvation energy relationships: octanol-water partition coefficients and aqueous solubilities. *J. Pharm. Sci.* **1986**, *75*, 629-636.

Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379-400.

Leo, A. J. Calculating $\log P_{oct}$ from structures. *Chem. Rev.* **1993**, *93*, 1281-1306.

Leo, A. J. Methods of calculating partition coefficients. In *Comprehensive Medicinal Chemistry*, vol. 4, Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamonn Press: New York, 1990; pp 295-319.

Leo, A. The octanol-water partition coefficient of aromatic solutes: the effect of electronic interactions, alkyl chains, hydrogen bonds, and ortho-substitution. *J. Chem. Soc. Perkin Trans. II* **1983**, 825-838.

Leo, A.; Hansch, C.; Elkins, D. (1971) Partition coefficients and their uses. *Chem. Rev.* **71**, 525-616.

Leo, A.; Jow, P. Y. C.; Silipo, C.; Hansch, C. Calculation of hydrophobic constant (log P) from π and f constants. *J. Med. Chem.* **1975**, *18*, 865-868.

Levin, V. A. Relationship of octanol/water partition coefficients and molecular weight to rat brain capillary permeability. *J. Med. Chem.* **1980**, *23*, 682-684.

Levine, I. N. *Physical Chemistry*, 4th edn., McGraw-Hill: New York, 1995; p 901.

Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. Improved methods for semiempirical solvation models. *J. Comput. Chem.* **1995**, *16*, 422-440.

Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting. *Adv. Drug Del. Rev.* **1997**, *23*, 3-25.

Livingstone, J. R.; Spolar, R. S.; Record, M. T., Jr. Contribution to the thermodynamics of protein folding from the reduction in water-accesible nonpolar surface area. *Biochemistry* **1991**, *30*, 4237-4244.

Lumry, R.; Rajender, S. Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers* **1970**, *9*, 1125-1227.

Mannhold, R.; Dross, K. P.; Rekker, R. F. Drug lipophilicity in qsar practice: I. A comparison of experimental with calculative approaches. *Quant. Struct.-Act. Relat.* **1990**, *9*, 21-28.

Mannhold, R.; Rekker, R. F.; Sonntag, C.; Laak, A. M.; Dross, K.; Polymeropoulos, E. E. Comparative evaluation of the predictive power of calculation procedures for molecular lipophilicity. *J. Pharm. Sci.* **1995**, *84*, 1410-1419.

Martindale The Extra Pharmacopoeia, 30th edn.; Reynolds, J. E. F., Ed.; The Pharmaceutical Press: London, 1993.

Masuda, T.; Jikihara, T.; Nakamura, K.; Kimura, A.; Takagi, T.; Fujiwara, H. Introduction of solvent-accessible surface area in the calculation of the hydrophobicity parameter log *P* from an atomistic approach. *J. Pharm. Sci.* **1997**, *86*, 57-63.

Meier, H. Zur Theorie der Alkoholnarkose. *Arch. Exp. Pathol. Pharmakol.* **1899**, *42*, 109-118.

Meot-Ner, M.; Sieck, L. W. The ionic hydrogen bond. 1. Sterically hindered bonds. solvation and clustering of protonated amines and pyridines. *J. Am. Chem. Soc.* **1983**, *105*, 2956-2961.

Merck Index, 12th edn.; Budavari, S., Ed.; Merck & Co. Inc: Whitehouse Station, NJ, 1996.

Meyer, A. Y. Molecular mechanics and molecular shape. III. Surface area and cross-sectional areas of organic molecules. *J. Comput. Chem.* **1986**, *7*, 144-152.

Meylan, W. M.; Howard, P. H. Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* **1995**, *84*, 83-92.

Miller, M. M.; Ghodbane, S.; Wasik, S. P.; Tewari, Y. B.; Martire, D. E. Aqueous solubilities, octanol/water partition coefficients, and entropies of melting of chlorinated benzenes and biphenyls. *J. Chem. Eng. Data* **1984**, *29*, 184-190.

Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127-130.

Moriguchi, I.; Kanada, Y.; Komatsu, K. Van der Waals volume and the related parameters for hydrophobicity in structure-activity studies. *Chem. Pharm. Bull.* **1976**, *24*, 1799-1806.

Morimoto, Y.; Hatanaka, T.; Sugibayashi, K.; Omiya, H. Prediction of skin permeability of drugs: Comparison of human and hairless rat skin. *J. Pharm. Pharmacol.* **1992**, *44*, 634-639.

Morita, H.; Kondo, K.; Hitotsuyanagi, Y.; Takeya, K.; Itokawa, H.; Tomioka, N.; Itai, A.; Iitaka, Y. Conformational analysis of antitumor cyclic hexapeptides, RA series. *Tetrahedron* **1991**, *47*, 2757-2772.

Motoc, I.; Marshall, R. Van der Waals volume fragmental constants. *Chem. Phys. Lett.* **1985**, *116*, 415-419.

Mullersman, G.; Derendorf, H.; Brewster, M. E.; Estes, K. S.; Bodor, N. High performance liquid chromatographic assay of a central nervous system (CNS)-directed estradiol chemical delivery system and its application after intravenous administration in rats. *Pharm. Res.* **1988**, *5*, 172-177.

Murray, J. S.; Ranganathan, S.; Politzer, P. Correlations between the solvent hydrogen bond acceptor parameter β and the calculated molecular electrostatic potential. *J. Org. Chem.* **1991**, *56*, 3734-3737.

Niemi, G. J.; Basak, S. C.; Veith, G. D.; Grunwald, G. Prediction of octanol/water partition coefficient (K_{ow}) with algorithmically derived variables. *Environ. Tox. Chem.* **1992**, *11*, 893-900.

Nyburg, S. C.; Faerman, C. H. A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br, and I bonded to carbon. *Acta Cryst.* **1985**, *B41*, 274-279.

Nys, G. G.; Rekker, R. F. Statistical analysis of a series of partition coefficients with special reference to the predictability of folding of drug molecules. The introduction of hydrophobic fragmental constants (f values). *J. Med. Chem.* **1973**, *8*, 521-535.

Nys, G. G.; Rekker, R. F. The concept of hydrophobic fragmental constants (f-values). II. Extension of its applicability to the calculation of lipophilicities of aromatic and heteroaromatic structures. *Eur. J. Med. Chem.* **1974**, *9*, 361-375.

Osinga, M. Rigidity, lipophilicity, and Henry's constant. *J. Am. Chem. Soc.* **1979**, *101*, 1621-1622.

Overton, E. *Studien über die Narkose*, Fischer: Jena, Germany, 1901.

Overton, E. Über die osmotischen Eigenschaften der Zelle in ihrer Bedeutung für die Toxikologie und Pharmakologie. *Z. Phys. Chem.* **1897**, *22*, 189-209.

Pardridge, W. M.; Triguero, D.; Yang, J.; Cancilla, P. A. Comparison of in vitro and in vivo models of drug transcytosis through the blood-brain barrier. *J. Pharmacol. Exp. Ther.* **1990**, *253*, 884-891.

Partington, J. R. *An Advanced Treatise on Physical Chemistry*, vol. 1., Fundamental Principles. The Properties of Gases. Longmans, Greene and Co.: London, 1949; p 943 (cited on pp. 664).

Pearlman, R. S. Molecular surface area and volume: their calculation and use in predicting solubilities and free energies of desolvation. In *Partition Coefficient Determination and Estimation,* Dunn, W. J., III., Block, J. H., Pearlman, R. S., Eds.; Pergamonn: New York, 1986; pp 3-20.

Perrin, D. D. *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworth: London, 1965.

Perrot, G.; Cheng, B.; Gibson, K. D.; Vila, J.; Palmer, K. A.; Nayeem, A.; Maigret, B.; Scheraga, H. A. MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives. *J. Comput. Chem.* **1992**, *13*, 1-11.

Petitjean, M. On the analytical calculation of van der Waals surfaces and volumes: some numerical aspects. *J. Comput. Chem.* **1994**, *15*, 507-523.

Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Möller, A.; Brickmann, J. Empirical method for the quantification and localization of molecular hydrophobicity. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309-1319.

Prokai, L.; Ouyang, X.-D.; Wu, W.-M.; Bodor, N. Chemical delivery system to transport a pyroglutamyl peptide to the central nervous system. *J. Am. Chem. Soc.* **1994**, *116*, 2643-2644.

Prokai, L.; Simay, A.; Bodor, N. Reversed-phase high-performance liquid chromatography of ketoxime analogues of β-adrenergic blockers. *J. Chromatogr.* **1991**, *541*, 469-473.

Prokai, L; Wu, W.-M.; Somogyi, G.; Bodor, N. Ocular delivery of the β-adrenergic antagonist alprenolol by sequential bioactivation of its methoxime analog. *J. Med. Chem.* **1995**, *38*, 2018-2020.

Prokai-Tatrai, K.; Prokai, L.; Bodor, N. Brain-targeted delivery of a leucine-enkephalin analogue by retrometabolic design. *J. Med. Chem.* **1996**, *39*, 4775-4782.

Rapoport, S. I.; Ohno, K.; Pettigrew, K. D. Drug entry into brain. *Brain. Res.* **1979**, *172*, 354-359.

Recanatini, M. Partition and distribution coefficients of aryloxypropanolamine β-adrenoceptor antagonists. *J. Pharm. Pharmacol.* **1992**, *44*, 68-70.

Reddy, I. K.; Bodor, N. In vitro evaluation of a controlled-release site-specific diisovaleryl tert-butalone chemical delivery system for the eye. *J. Pharm. Sci.* **1994**, *83*, 450-453.

Rekker, R. F. *The Hydrophobic Fragmental Constant*, Elsevier: New York, 1977; p 389.

Rekker, R. F.; de Kort, H. M. The hydrophobic fragmental constant: an extension to a 1000 data point set. *Eur. J. Med. Chem.* **1979**, *14*, 479-488.

Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity*. VCH Publishers: Weinheim, 1992.

Richards, F. M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151-176.

Richards, N. G. J.; Williams, P. B.; Tute, M. Empirical methods for computing molecular partition coefficients. I. Upon the need to model the specific

hydration of polar groups in fragment-based approaches. *Intern. J. Quantum Chem.: Quantum Biol. Symp.* **1991**, *18*, 299-316.

Richards, N. G. J.; Williams, P. B.; Tute, M. Empirical methods for computing molecular partition coefficients. II. Inclusion of conformational flexibility within fragment-based approaches. *Intern. J. Quantum Chem.* **1992**, *44*, 219-233.

Richmond, T. J. Solvent accessible surface area and excluded volume in proteins. *J. Mol. Biol.* **1984**, *178*, 63-89.

Rogers, K. S.; Cammarata, A. A molecular orbital description of the partitioning of aromatic compounds between polar and nonpolar phases. *Biochim. Biophys. Acta* **1969a**, *193*, 22-29.

Rogers, K. S.; Cammarata, A. Superdelocalizability and charge density. A correlation with partition coefficients. *J. Med. Chem.* **1969b**, *12*, 692-693.

Rosenberg, S. H.; Woods, K. W.; Sham, H. L.; Kleinert, H. D.; Martin, D. L.; Stein, H.; Cohen, J.; Egan, D. A.; Bopp, B.; Merits, I.; Garren, K. W.; Hoffman, D. J.; Plattner, J. J. Water-soluble renin inhibitors: design of subnanomolar inhibitor with a prolonged duration of action. *J. Med. Chem.* **1990**, *33*, 1962-1969.

Rowlinson, J. S. The triplet distribution function in a fluid of hard spheres. *Mol. Phys.* **1963**, *6*, 517-524.

Samii, A.; Bickel, U.; Stroth, U.; Pardridge, W. M. Blood-brain barrier transport of neuropeptides: analysis with a metabolically stable dermorphin analogue. *Am. J. Physiol.* **1994**, *267*(1 Pt 1), E124-E131.

Sangster, J. Octanol-water partition coefficients of simple organic compounds. *J. Phys. Chem. Ref. Data* **1989**, *18*, 1111-1229.

Sarraf, E. G. New thermodynamic model for the estimation of partition coefficients. Part 1. Weakly hydrogen-bonded solutes. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 2519-2525.

Schoenwald, R. D.; Huang, H.-S. Corneal penetration behavior of β-blocking agents I: Physicochemical factors. *J. Pharm. Sci.* **1983**, *72*, 1266-1272.

Seydel, J. K.; Schaper, K. J. Quantitative structure-pharmacokinetic relationships and drug design. *Pharmac. Ther.* **1982**, *15*, 131-181.

Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting hydrophobic free energies from experimental data: Relationship to protein folding and theorethical models. *Biochemistry*, **1991**, *30*, 9686-9697.

Shim, C. K.; Nishigaki, R.; Iga, T.; Hanano, M. Determination of extraction constant, true partition coefficient and formation constant of ion-pair complexes of quaternary ammonium salts, tetrabutylammonium bromide and isopropamide iodide, with some organic anions by a solvent extraction technique. *Int. J. Pharm.* **1981**, *8*, 143-151.

Shirley, B. A.; Stanssens, P.; Hahn, U.; Pace, C. N. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **1992**, *31*, 725-732.

Shiu, W. Y.; Mackay, D. A critical review of aqueous solubilities, vapor pressures, henry's law constants, and octanol-water partition coefficients of the polychlorinated biphenyls. *J. Phys. Chem. Ref. Data* **1986**, *15*, 911-929.

Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozime and insulin. *J. Mol. Biol.* **1973**, *79*, 351-371.

Silla, E.; Tunon, I.; Pascual-Ahuir, J. L. GEPOL: An improved description of molecular surfaces. II. Computing the molecular area and volume. *J. Comput. Chem.* **1991**, *12*, 1077-1088.

Silla, E.; Tunon, I.; Villar, F.; Pascual-Ahuir, J. L. Molecular surface calculations on organic compounds. Molecular area - aqueous solubility relationships. *Theochem (J. Mol. Struct.)* **1992**, *254*, 369-377.

Simay, A.; Prokai, L.; Bodor, N. Oxidation of aryloxy-β-amino alcohols with activated dimethylsulfoxide. A novel C-N oxidation facilitated by neighboring group effect. *Tetrahedron* **1989**, *45*, 4091-4102.

Stein, W. D. *Transport and Diffusion across Cell Membranes*, Academic Press: Orlando, 1986; p 685.

Suzuki, T.; Kudo, Y. Automated logP estimation based on combined additive modeling methods. *J. Comput-Aided Mol. Des.* **1990**, *4*, 155-198.

Taft, R. W.; Abraham, M. H.; Doherty, R. M.; Kamlet, M. J. The molecular properties governing solubilities of organic nonelectrolytes in water. *Nature* **1985a**, *313*, 384-386.

Taft, R. W.; Abraham, M. H.; Famimi, G. R.; Doherty, R. M.; Abboud, J.-L. M.; Kamlet, M. J. Solubility properties in polymers and biological media. 5. An analysis of the physicochemical properties which influence octanol-water partition. *J. Pharm. Sci.* **1985b**, *74*, 807-814.

Taft, R. W.; Berthelot, M.; Laurence, C.; Leo, A. J. Hydrogen bonds and molecular structure. *Chemtech* **1996**, *26 (7)*, 20-29.

Takada, K.; Usuda, H.; Oh-Hashi, M. Distribution kinetics of FK-506, a novel immunosuppresant, after intravenous administration to rats in comparison with cyclosporin A. *Biopharm-Drug-Dispos.* **1992**, *13*, 345-355.

Taylor, P. J. Hydrophobic properties of drugs. In *Comprehensive Medicinal Chemistry*, vol. 4, Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: New York, 1990; pp 241-294.

Timmermans, P. B. M. W. M.; Brands, A.; van Zwieten, P. A. Lipophilicity and brain disposition of clonidine and structurally related imidazolidines. *Naunyn-Schmiedberg's Arch. Pharmacol.* **1977**, *300*, 217-226.

Testa, B.; Carrupt, P.-A.; Gaillard, P.; Billois, F.; Weber, P. Lipophilicity in molecular modeling. *Pharm. Res.* **1996**, *13*, 335-343.

Tewari, Y. B.; Miller, M. M.; Wasik, S. P.; Martire, D. E. *J. Chem. Eng. Data* **1982**, *27*, 451-454.

Tokumura, T.; Tanaka, T.; Sasaki, A.; Tsuchiya, Y.; Abe, K.; Machida, R. Stability of a novel hexapeptide, (Me)Arg-Lys-Pro-Trp-tert-Leu-Leu-OEt, with neurotensin activity, in aqueous solution and in the solid state. *Chem. Pharm. Bull.* **1990**, *38*, 3094-3098.

Tomida, H.; Yotsuyanagi, T.; Ikeda, K. Solubilization of steroid hormones by polyoxyethylene lauryl ether. *Chem. Pharm. Bull.* **1978**, *26*, 2832-2837.

Tomlinson, E. Enthalpy-entropy compensation analysis of pharmaceutical, biochemical, and biological systems. *Int. J. Pharm.* **1983**, *13*, 115-144.

Tsuji, A.; Tamai, I. Blood-brain barrier function of p-glycoprotein. *Adv. Drug Del. Rev.* **1997**, *25*, 287-298.

Tsuji, A.; Tamai, I.; Sakata, A.; Tenda, Y.; Terasaki, T. Restricted transport of cyclosporin A across the blood-brain barrier by a multidrug transporter, P-glycoprotein. *Biochem. Pharmacol.* **1993**, *46*, 1096-1099.

Valvani, S. C.; Yalkowsky, S. H.; Amidon, G. L. Solubility of nonelectrolytes in polar solvents. VI. Refinements in molecular surface computation. *J. Phys. Chem.* **1976**, *80*, 829-835.

Van Asperen, J.; Mayer, U.; Van Tellingen, O.; Beijnen, J. H. The functional role of P-glycoprotein in the blood-brain barrier. *J. Pharm. Sci.* **1997**, *86*, 881-884.

Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an

automated superposition of certain naturally occuring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163-172.

Viswanadhan, V. N.; Reddy, M. R.; Bacquet, R. J.; Erion, M. D. Assessment of methods used for predicting lipophilicity: Application to nucleosides and nucleoside bases. *J. Comput. Chem.* **1993**, *14*, 1019-1026.

Wang, R.; Fu, Y.; Lai, L. A new atom-additive method for calculating partition coefficients. *J. Chem. Inf. Comp. Sci.* **1997**, *37*, 615-621.

Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **1996**, *35*, 5109-5124.

Yalkowsky, S. H.; Pinal, R.; Banerjee, S. Water solubility: A critique of the solvatochromic approach. *J. Pharm. Sci.* **1988**, *77*, 74-77.

Young, R. C.; Mitchell, R. C.; Brown, T. H.; Ganellin, C. R.; Griffiths, R.; Jones, M.; Rana, K. K.; Saunders, D.; Smith, I. R.; Sore, N. E.; Wilks, T. J. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H_2 receptor histamine antagonists. *J. Med. Chem.* **1988**, *31*, 656-671.

Yunger, L.M., Cramer, R.D., III. Measurement and correlation of partition coefficients of polar amino acids. *Mol. Pharmacol.* **1981**, *20*, 602-608.

BIOGRAPHICAL SKETCH

Peter Buchwald was born on February 27, 1963, in Cluj (Kolozsvár), Romania. He completed his school education in the same city. In 1982, he entered the Babes-Bolyai University of Cluj (Kolozsvár), Romania, and graduated in 1986 with a B.S. in physics. In 1992, after working three years as a teacher and two years as computerized information manager, he joined the graduate program of the Center for Drug Discovery, College of Pharmacy, University of Florida, and completed his Ph.D. in 1997.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Nicholas S. Bodor, Chair
Graduate Research Professor of
Pharmaceutics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Hartmut Derendorf
Professor of Pharmaceutics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William J. Millard
Professor of Pharmacodynamics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

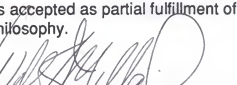
Laszlo Prokai
Associate Professor of
Pharmaceutics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David A. Micha
Professor of Chemistry and Physics

This dissertation was submitted to the Graduate Faculty of the College of Pharmacy and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1997



Dean, College of Pharmacy

Dean, Graduate School